THE PROCEEDINGS OF
THE PHYSICAL SOCIETY

Section B

VOL. 62, PART 7 1 July 1949 No. 355 B

Multiple-Beam Interferometry: Intensity Distribution in the Reflected System

BY J. HOLDEN

Royal Holloway College, University of London, Englefield Green, Surrey

Communicated by S. Tolansky; MS. received 29th October 1948

ABSTRACT. The intensity distribution within the reflected fringe system from a multiple-beam interferometer employing silvered reflecting surfaces is discussed in its dependence upon the reflectivities and phase conditions at the reflecting surfaces. Using the Fizeau fringes of equal thickness localized on the zero order Feussner surface, the reflected intensity distribution is examined experimentally as the reflection coefficients of the interferometer surfaces vary between 4% and 90%. In the range of low reflectivities, when the silverings used showed marked colours, the changes in the reflected system lead to the measurement of a phase quantity related to the optical constants of the silver in the form of the thin film. Also in this range, the reflected fringes have a symmetrical form which is of use in the examination of sources of low intensity. In the range of high reflectivities, the conditions for the use of the reflected fringes in topographical investigation are discussed. The findings apply to fringes of equal chromatic order as well as to Fizeau fringes in reflection.

§ 1. INTRODUCTION

THE reflected interference fringe systems find many applications in metrology, the "Fizeau" localized fringes of equal thickness being most frequently used. The surfaces of the interferometer usually carry a thin, partially reflecting layer of silver, but a general description of the reflected intensity distribution from such reflectors for the particular reflecting coefficients involved appears to be confined to a little-known early paper by Hamy (1906), who used chemically deposited silver to form the reflecting surfaces.

In the work to be described here, an investigation is made of the reflected intensity distribution given by an interferometer whose reflecting surfaces are formed by the thermal deposition of silver upon optical flats, so as to cover the range of reflectivities from 4% to 92%. Experimentally, the Fizeau fringes localized on the zero order Feussner surface are used for the observations, but the results obtained also apply to fringes of equal inclination and to fringes of equal chromatic order (Tolansky 1945) in reflection.

The observations of Hamy have been confirmed and extended. In the range of low reflectivities, when the reflecting silver films are less than 100 A. thick, the relation of the measurable phase quantities associated with the reflected fringes to the more usual measurements of phase quantities in single films is discussed. In the range of high reflectivities the conditions for the use of reflected fringes in topographical measurements, such as have been carried out using the transmitted system in the case of a diamond surface (Tolansky and Wilcock 1947), are derived.

§ 2. THE INTENSITY REFLECTED FROM A PLANE PARALLEL SHEET

The case of a plane parallel air film between silvered optical flats will first be considered. The properties of such a system as developed theoretically can be tested experimentally using a simple wedge interferometer if the wedge angle and separation are sufficiently small (Brossel 1947).

Figure 1 illustrates two such reflecting planes, with normally incident, monochromatic, parallel light giving rise to families of reflected and transmitted wave fronts.

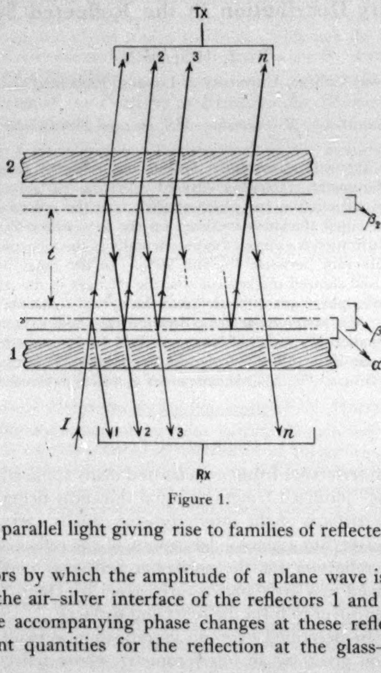


Figure 1.

Let the factors by which the amplitude of a plane wave is caused to change on reflection at the air–silver interface of the reflectors 1 and 2 be r_1 and r_2 and let β_1, β_2 be the accompanying phase changes at these reflections. Let r_α, α be the equivalent quantities for the reflection at the glass–silver interface of reflector 1.

Finally, let t_1 and t_2 be the amplitude transmission factors of the surfaces 1 and 2 and τ_1, τ_2 the phase changes at these transmissions. The changes of phase β_1, β_2, α are illustrated in a geometrical manner in the figure, that is, by assuming them equivalent to a change in air gap, of arbitrary sign.

Assuming a plane wave of unit amplitude to be incident normally upon the interferometer, the state of interference at any point X in the focal plane of a collecting lens is given by the vector sum of the disturbances due to each reflected wave. For the reflected system the resulting vibration at X is given by:

$$v_X = r_\alpha \cos(\theta - \alpha) + \sum_{n=1} t_1^2 r_2^n r_1^{n-1} \cos\{[\theta - 2\tau_1] - [(n-1)\beta_1 + n(\beta_2 + \delta)]\}.$$

Assuming that the absorption is the same for each direction of transit through a film then the reflected intensity is

$$I_{\mathrm{RX}} = r_\alpha^2 + \left\{ \frac{(t_1^2 r_2)^2 + 2t_1^2 r_2 r_\alpha \cos(\delta - \alpha + \beta_2 + 2\tau_1) - 2t_1^2 r_2^2 r_1 r_\alpha \cos(2\tau_1 - \beta_1 - \alpha}{1 + r_1^2 r_2^2 - 2r_1 r_2 \cos(\delta + \beta_1 + \beta_2)} \right\}.$$

In the same nomenclature the transmitted intensity is

$$I_{\mathrm{TX}} = \left\{ \frac{t_1^2 t_2^2}{1 + r_1^2 r_2^2 - 2r_1 r_2 \cos(\delta + \beta_1 + \beta_2)} \right\} = \Theta_\Delta [t_1^2 t_2^2],$$

writing

$$\left\{ \frac{1}{1 + r_1^2 r_2^2 - 2r_1 r_2 \cos(\delta + \beta_1 + \beta_2)} \right\} = \Theta_\Delta,$$

where $\Delta = \delta + \beta_1 + \beta_2$, the optical separation of the two surfaces. This term is familiar in its rôle in the Fabry–Perot interferometer, which makes use of the rapid variation of Θ_Δ about $\delta + \beta_1 + \beta_2 = \Delta = 2n\pi$ as $r_1 r_2 \to 1$.

Since $(2\tau_1 - \beta_1 - \alpha) = F$ depends only on the phase conditions at the first surface, 1, the reflected system now becomes

$$I_{\mathrm{RX}} = r_\alpha^2 + \Theta_\Delta [(t_1^2 r_2)^2 + 2t_1^2 r_2 r_\alpha \cos(\Delta + F) - 2t_1^2 r_2^2 r_1 r_\alpha \cos F].$$

For the simplified case of a symmetrical interferometer where $r_1^2 = r_2^2 = R$ and $t_1^2 = t_2^2 = T$, the expressions become

$$I_{\mathrm{TX}} = \Theta_\Delta [T^2],$$

$$I_{\mathrm{RX}} = R + \Theta_\Delta [T^2 R + 2TR \cos(\Delta + F) - 2TR^2 \cos F].$$

It is seen that while the transmitted system is described by an expression consisting of the function Θ_Δ (the Airy "shape" function), multiplied by a magnitude term involving only the transmission factors of the two surfaces, the reflected system differs since in addition to a superposed background term R there are two transmission-like terms, one of which involves $\cos F$ in its magnitude; there is also a term which is unique, since it is out of phase with Θ_Δ by an amount F.

The condition leading to maxima and minima in the expression for the reflected intensity I_{RX} is

$$\sin \Delta \{ t_1^2 r_2^2 r_1 r_2 + r_2 r_\alpha \cos F - r_2^2 r_1^2 r_2 r_\alpha \cos F \}$$
$$+ \cos \Delta \{ r_2 r_\alpha \sin F + r_1^2 r_2^2 r_2 r_\alpha \sin F \} = 2r_2^2 r_1 r_\alpha \sin F,$$

which can be written

$$m \sin \Delta + n \cos \Delta = p,$$

where m, n and p are functions of r_1, r_2, r_α, t_1 and F. The solutions of this equation can be written

$$\Delta = 2n\pi + a; \qquad \Delta = (2n+1)\pi - b,$$

where one of the series of roots corresponds to maxima and the other to minima, the quantities a and b being of the order of π and dependent upon the values of r_1, r_2, r_α, t_1 and F, that is, upon the reflection coefficients of the two surfaces of the interferometer, the transmission coefficient and the phase changes at the first (incident) surface. The two series of roots will not, in general, occur at equal intervals.

The conditions that the roots be equally spaced is that in the equation $\partial I_{\mathrm{RX}} / \partial \Delta = 0$ the independent term p is zero, i.e. $2r_2^2 r_1 r_\alpha \sin F = 0$, which can occur when $\sin F=0$. Two cases arise, namely, (i) $F=2n\pi$, and (ii) $F=(2n+1)\pi$. In the first case, intensity maxima occur at $\Delta=2n\pi$ and minima at $\Delta=(2n+1)\pi$. In the second case minima occur at $\Delta=2n\pi$ and maxima at $\Delta=(2n+1)\pi$.

The other limiting condition of the intensity distribution which arises from the form of the equation $\partial I_{\mathrm{Rx}}/\partial\Delta=0$ occurs when the two roots are coincident, that is when (iii) $m^2+n^2=p^2$ and

$$F=(2n+1)\pi; \quad t_1^2=\frac{r_2 r_\alpha(1-r_1^2 r_2^2)}{r_2^2 r_1 r_2}.$$

In the absence of any known relations between the phase quantities α, β_1, τ_1 for a particular silver film it is seen that the reflected fringe system from an interferometer in which the silver film is the incident surface may be expected to be asymmetrical. Observation of this asymmetry with relation to the symmetrical transmitted fringe system from the same interferometer will allow the empirical determination of F for different film thicknesses. Of the three limiting conditions when the reflected fringe system is symmetrical, it is found experimentally for films of silver upon glass that cases (i) and (iii) represent limits to which actual conditions can tend, while case (ii) arises when the glass is unsilvered.

§ 3. EXTENSION TO THE CASE OF A WEDGE

It has been shown in the case of a silvered wedge interferometer (Brossel 1947) that the phase of the nth transmitted beam with respect to the direct is closely given by

$$\phi_n=n\delta-n^3 8\pi t\epsilon^2/3\lambda,$$

as compared with $\phi_n=n\delta$ for strictly parallel surfaces, where ϵ is the wedge angle and t the wedge separation at the point considered. It is found experimentally that if the terms $n^3 8\pi t\epsilon^2/3\lambda$ are small the transmitted fringes are asymmetrical, and as the terms increase the asymmetry resolves into secondary maxima close to the main maxima. The reflected fringes from a wedge also show asymmetry due to this effect but, in order to obviate confusion with the asymmetry due to the intrinsic phase characteristics of surface 1, it can be made negligible by choosing an appropriate wedge angle and order of interference to make the phase lag terms due to the presence of the wedge angle negligibly small. The deductions made in the case of a parallel sheet thus apply to a wedge chosen in this way.

§ 4. EXPERIMENTAL

The optical arrangement used consists of a source arm complete with collimating system which can be moved with respect to a fixed wedge interferometer and viewing system, so that the transmitted and reflected fringe systems localized on the Feussner surface of zero order can be observed in turn.

Special attention must be given to the collimation and the angle of incidence of the light in each of the two positions. It is found experimentally that the reflected intensity distributions close to the Feussner surface of zero order F_0 have a markedly asymmetrical appearance with respect to the appearance of the distribution on the Feussner surface. Focusing further away from the Feussner surface causes the asymmetry quickly to attain the appearance of an adjacent maximum and minimum against the general background. Figure 2 shows the appearance of the reflected system (below) when the plane of focus is relatively steeply inclined to the Feussner surface and intersecting it. The accompanying transmitted system photographed on the same plate (above) by means of a split

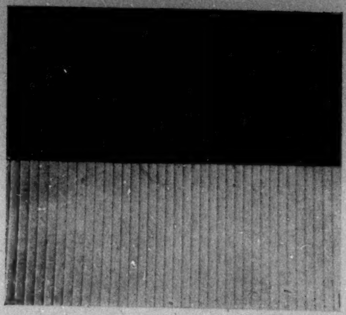


Figure 2.

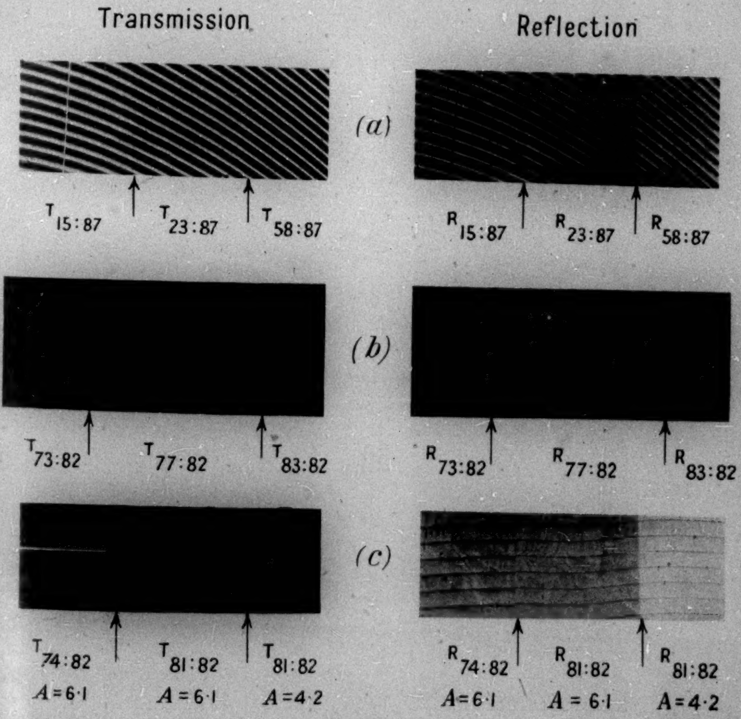

Transmission
Reflection
(a)
$^{T}15:87$
$^{T}23:87$
$^{T}58:87$
$^{R}15:87$
$^{R}23:87$
$^{R}58:87$
(b)
$^{T}73:82$
$^{T}77:82$
$^{T}83:82$
$^{R}73:82$
$^{R}77:82$
$^{R}83:82$
(c)
$^{T}74:82$ $A=6{\cdot}1$
$^{T}81:82$ $A=6{\cdot}1$
$^{T}81:82$ $A=4{\cdot}2$
$^{R}74:82$ $A=6{\cdot}1$
$^{R}81:82$ $A=6{\cdot}1$
$^{R}81:82$ $A=4{\cdot}2$

Figure 5.

To face page 408

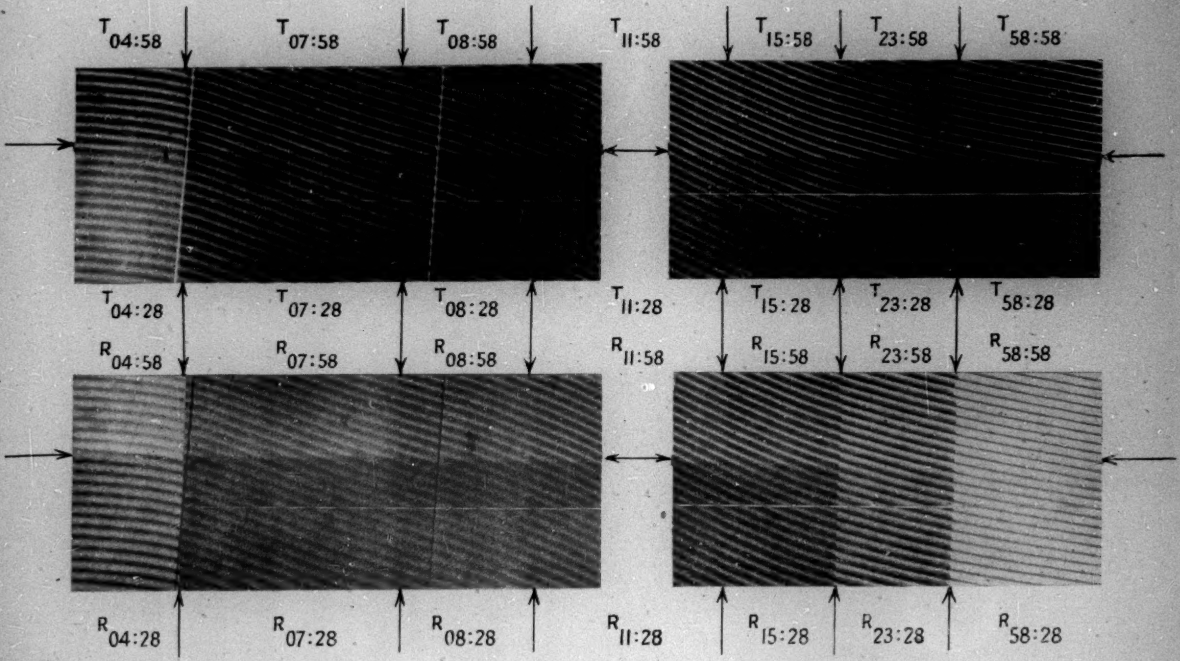

$T_{04:58}$
$T_{07:58}$
$T_{08:58}$
$T_{11:58}$
$T_{15:58}$
$T_{23:58}$
$T_{58:58}$
$T_{04:28}$
$T_{07:28}$
$T_{08:28}$
$T_{11:28}$
$T_{15:28}$
$T_{23:28}$
$T_{58:28}$
$R_{04:58}$
$R_{07:58}$
$R_{08:58}$
$R_{11:58}$
$R_{15:58}$
$R_{23:58}$
$R_{58:58}$
$R_{04:28}$
$R_{07:28}$
$R_{08:28}$
$R_{11:28}$
$R_{15:28}$
$R_{23:28}$
$R_{58:28}$

Figure 4.

shutter camera, under the same conditions, does not reveal its asymmetry at normal exposures. Over a small field it is clear that this asymmetry will be confusing unless it can be made negligibly small. The distribution close to the surface F_0 has been studied (Brossel 1947) in transmission, and it obeys the phase relation $\phi_n = n\delta - n^2 4\pi X\epsilon^2/\lambda$, where ϕ_n is the phase of the nth beam relative to the direct at the point X, and where X is the distance in the direction normal to the wedge from the common line of the wedge of angle ϵ. This relation has the same general properties as that described when considering the effects of wedge angle; it is therefore important to have X and ϵ such that the terms $n^2 4\pi X\epsilon^2/\lambda$ are negligibly small. If the incident light is collimated so that the angles of incidence are between the normal and a maximum of θ, the position of the Feussner surface $[F_0]_\theta$ for the light incident at θ is (Figure 3) $X_\theta = (t/\epsilon)\sin\theta$, where X_θ is the distance normal to the interferometer at a thickness t. If the plane of focus coincides with the Feussner surface $[F_0]_0$ for the normally incident light, it will be distant an amount $\{(t/\epsilon)\sin\theta\}$ from the surface $[F_0]_\theta$, which is the Feussner surface for the light incident at an angle θ. Therefore θ must be chosen such that the terms $n^2 4\pi X_\theta\epsilon^2/\lambda$ are small compared with $n\delta$.

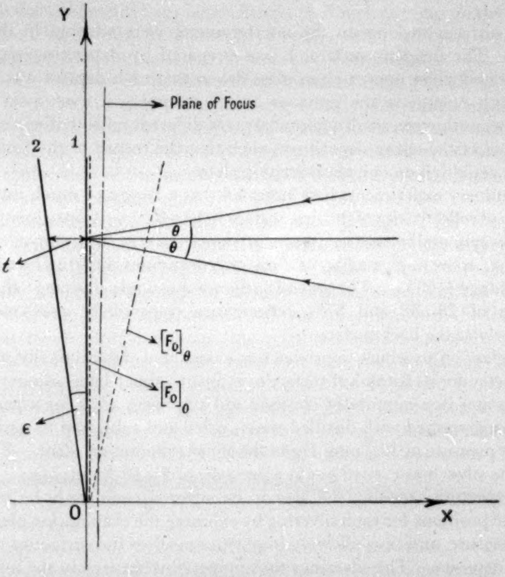


Figure 3.

Another reason why the collimation should be especially rigorous is that the incident beam has to illuminate the interferometer from two successive positions such that the transmission surface $[F_0]_{TX}$ and the reflection surface $[F_0]_{RX}$ are coincident within the same toleration as that discussed above for the Feussner surfaces within the one system. By using a collimating lens of focal length

20 cm. and a source aperture 1 mm. in diameter, θ is reduced to about 15′. Using a wedge angle of less than 1′, and working with an order of interference of less than 25, the focusing and wedge angle terms can be made negligible.

Finally, the lens used for the observations must be of appropriate aperture to collect sufficient of the multiply reflected emergent beams. The lens chosen was of focal length 50 mm., f/3. Figure 2 (see Plate) shows the effects of using an inadequate lens. Apart from the curvature of the field, the point of interest interferometrically is that there exists a variation of fringe shape along the length of the fringe because the aperture of the lens is too small for the wedge angle used. In the figure are shown the transmitted (above) and reflected (below) fringes from the wedge; these errors are much more obvious in the reflected system where the light intensity is high and the background level between the fringes not close to zero as in the transmitted system where excessive over-exposure is needed to reveal them.

The surfaces making up the interferometer were silvered in the following manner. The incident surface, 1, was prepared by depositing six thin layers of evaporated silver upon a clean glass flat so that each deposit was slightly displaced with respect to the previous one, overlapping it over most of its area. By this means thirteen small adjacent areas of different reflectivities were obtained. The extent of the entire deposit was such that the record of the fringes from all the areas could fall on one photographic plate.

Preliminary experiments had indicated that a region of much interest was in the range of reflectivities of the first surface below 40%. Consequently the layers of silver evaporated in each successive process were very thin, equivalent thickness about 30 A., so as to give adjacent areas of reflectivities differing by a few per cent over the range 4–60%. The second surface was a single silvering; three different silverings of 28, 58 and 87% reflectivities respectively were made for use successively as the back surface.

The glass on to which the silver was evaporated was originally photographic plate selected for its flatness of surface over an area about 1 cm. square, and cleaned by immersion in a mixture of chromic and sulphuric acids for a period of days before being washed with distilled water, dried and subjected to ionic bombardment at a pressure of 10^{-2} mm. Hg in the silver evaporation plant. The evaporation of the silver was carried out at a pressure of 5×10^{-4} mm. Hg.

The aperture exposing the area to be silvered could only be moved to its successive positions for each silvering by exposing the evaporation plant to atmospheric pressure, and thus allowing impurities to affect the surface of the silvering between deposits. The silverings thus prepared were used in the interferometer within a few hours of their preparation, and no inconsistency in the observations due to possible changes in the films in this period was apparent. The thin silverings with reflection coefficients between 4 and 20% exhibited a striking range of colour both in transmission and in reflection, particularly from the glass–silver reflection, if the accompanying air–glass reflection was avoided. The colour changes were fully appreciated when the films of different thickness were matched side by side as in the case of these compound silvered interferometer surfaces. On one typical test surface six areas of progressively increasing reflectivity fell in the range below 20% reflectivity, and the film colours in transmission were clear yellow, yellow-orange, orange, purple, dark purple, blue. The films of reflectivity greater than 20% showed the characteristic blue transmission colour of silver films.

§5. OBSERVATIONS AND MEASUREMENTS OF THE REFLECTIVITIES AND THE PHASE QUANTITY F

Figure 4 shows the localized Fizeau fringes from a wedge interferometer whose incident surface was formed of six silverings to give areas of reflectivities 4, 7, 8, 11, 15, 23 and 58%, and whose back surface was formed of two silverings of reflectivities 28 and 58%, the field of view being arranged so that the seven areas on the front surface covered each of the different back surface areas. In the Figure the reflected and transmitted systems are indicated by R and T respectively, with appropriate subscripts to denote the reflectivities of the front and back surfaces for the particular area, the front surface being given first. In this particular compound silvering the final evaporation was rather heavy, bringing the maximum reflectivity of the area upon which it was deposited to 58%; the second sequence of reflectivities on the surface was therefore made in the reflectivity range 35 to 58%, which does not exhibit any large changes in F such as occur on the range 4 to 25%, and the fringes from these areas have been omitted in Figure 4.

The figure shows the variation in intensity distribution of the fringes as the reflection coefficient of the first surface increases from 4 to 58%. With the aid of the transmitted fringes, which have gradually sharpening maxima as the reflectivities increase from left to right, and of the two vertical and one horizontal lines which are the images of fine scratches ruled on the silver surface, it is easy to see the relative shift of the maxima and minima in the reflected system. When the first surface is uncoated glass ($r_1^2=0{\cdot}04$, $F=(2n+1)\pi$) the reflected system shows minima coincident with the transmitted maxima. (The area $R_{04:58}$ is of some incidental interest since then the reflectors have greatly disparate reflectivities and yet the fringes are easily visible.) As the reflectivity of the first surface increases, the fringes go through an asymmetrical sequence until at about 15% the reflected system shows sharp maxima almost coincident with the transmission maxima but much more luminous (the relative times of exposure of the two systems was 5 : 1). These fringes were first observed by Hamy (1906) and, noting their sharpness and high light efficiency, he suggested their use in the examination of weak sources. He gave only a verbal description of the fringes and little attention seems to have been given to his experimental findings. The sharpness attainable by the "transmission-like" fringes is shown in Figure 5 (*a*), area $R_{15:87}$, wherein the back surface is made to have a reflectivity of almost 90% and the light losses are due only to the double transit through the 15% reflecting silver film.

As the reflectivity on the first surface is further increased, the fringes again become asymmetrical, the background intensity grows, and finally the visual appearance is that of fine minima. Above 65% a microphotometer is needed to detect the asymmetry, and above 85% this asymmetry due to the phase relations at the first surface can no longer be certainly differentiated from that of other sources such as the wedge angle effect and the focusing defects.

Since in the illustrations the reflected fringes were in each case recorded on the same plate as were the transmitted fringes, the relative intensity levels within the systems from each area are roughly comparable in Figure 4. Figure 5 (*a*) shows fringes from an interferometer made up of the same front surface as the upper right plate in Figure 4, but with a back surface of 87% reflectivity. It is seen that the only effect of increasing the reflectivity of the back surface is to sharpen up all the fringes of the sequence while they retain their relative shapes, which are derived from the front surface conditions.

The transmitted and reflected systems shown in Figure 4 were microphotometered using a photographic recording instrument. Using the known plate response characteristic, the true fringe profiles were obtained, from which values of Δ satisfying the equation $\partial I_{\mathrm{RX}}/\partial\Delta = 0$ were measured. Substitution of these values in the equation gives values of $F(=2\tau_1 - \beta_1 - \alpha)$. These are plotted in Figure 6 against the reflection coefficients for the air–silver reflections at the front surfaces. These reflectivities were measured by a Strong type reflectometer and confirmed by a graphical method from the transmission fringe shape, obtained by using the microphotometer in conjunction with the equation expressing the intensity distribution in the transmitted system.

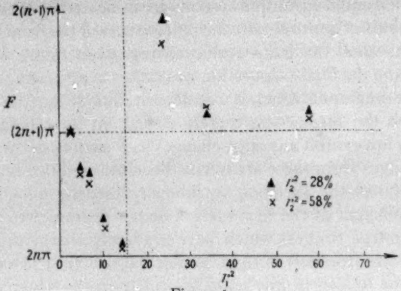


Figure 6.

§ 6. THE REFLECTED SYSTEM AT LOW REFLECTION COEFFICIENTS

The points in Figure 6 show a jump of 2π in the value of F in the region of 15 to 20% reflectivity of the silver film. In the many reported investigations of the phase quantities α and β of thin silver films some observers (Rouard 1937) have noticed a jump in the values of α at very small thicknesses, but the experimental difficulties (poor sharpness and visibility of Fabry–Perot fringes in transmission at the low reflectivities of the thin films involved) are considerable. Observation of the reflected fringe system from an interferometer, the front (incident) surface of which is the silver film under examination, immediately reveals the variation of the phase quantity F. From the discussion of symmetry of the reflected fringes, this experimental observation, that the fringes change from symmetrical minima when the front surface is unsilvered through an asymmetrical minimum–maximum form until they are almost symmetrical maxima when the surface is silvered to about 15%, means that F has changed from $(2n+1)\pi$ to $2n\pi$. As the reflectivity further increases, the fringes first have an asymmetrical maximum–minimum form and then a symmetrical minimum as the reflectivity increases to 60%, the minimum becoming sharper as the very highest reflectivities, 90%, are approached. This implies that F is tending to $(2n+1)\pi$, approaching it, however, from values between $(2n+1)\pi$ and $2(n+1)\pi$ as the reversed sense of the asymmetry requires. Figure 4 shows the sequence for the wavelength 5461 A.

The rapid variation of F in the range of very low reflectivities, and the colours which the metal films present, appear to be related to the state of aggregation of the metal of the film, since these phenomena occur for equivalent film thicknesses which are very small for explanations based on interference. The measurement of F can be extended to all visible wavelengths if the fringes of equal chromatic order (Tolansky 1945) are observed with the interferometer.

The relatively sharp "transmission-like" reflection fringes (cf. Figure 5 (*a*), area $\mathrm{R}_{15:87}$), appear to be of value if sources of low intensity only are available, since the reflected light only passes through a 15% reflecting silvering twice, and is thus not seriously reduced in intensity. These symmetrical fringes occur when $F=2n\pi$, which condition, in the case of the wavelength 5461 A. and air–silver–glass, arises when the reflectivity air–silver is about 15% (equivalent film thickness approximately 50 A.). In the case of other metal films and other dielectrics this condition, $F=2n\pi$, should occur at characteristic values of the film thickness. This may mean that a greater reflectivity is possible on the incident surface of an interferometer, the reflected fringes still appearing as symmetrical maxima on a dark background. In the case of the system air–silver–glass the "transmission-like" fringes have a half-width which is $\frac{1}{5}$ the distance between orders, when the back surface of the interferometer is opaque (for the green region). These fringes are also of use for the interferometric study of vibrating metallic surfaces.

§ 7. THE REFLECTED SYSTEM AT HIGH REFLECTION COEFFICIENTS

When the reflection coefficient of the first surface exceeds 60%, the reflected fringes present the visual appearance of fine minima on a background (Figure 5 (*b*)), and a microphotometer is needed to detect the slight asymmetry still present. The present measurement of F in this region is not capable of high accuracy since the quantities to be measured are then small (shifts of $\Delta_{\max}$ and $\Delta_{\min}$) and in the equation for the condition $\partial I_{\mathrm{RX}}/\partial\Delta=0$; as $r_1^2\rightarrow 1$, $r_2^2\rightarrow 1$, small errors in the measured quantities cause large variations in the calculated values of F. In this range F can only be measured to an accuracy of $\lambda/30$, and for $0{\cdot}6<R_1<1{\cdot}0$ is found to be equal to $(2n+1)\pi\pm\pi/10$.

From the earlier discussion it was seen that if the two conditions $F\rightarrow(2n+1)\pi$, $R_1\rightarrow 1$ could occur together the maxima and minima in the reflected intensity distribution would tend to be superimposed. It is observed experimentally that the reflected fringes (sharp minima) at high reflection coefficients become sharper than the accompanying transmitted system, whose sharpness is controlled by the function Θ_Δ alone. It is thus apparent that in the reflected system the superposition effect consequent upon $F\rightarrow(2n+1)\pi$ is taking place. For the present investigation in the region of high reflectivities the above accuracy of measurement of F is sufficient.

In order to examine the sharpness and contrast of the reflection fringes in the range $0{\cdot}6<R_1<1{\cdot}0$, which is the working range in topographical studies, we can take with good approximation $F=(2n+1)\pi$. Therefore the reflected intensity from a symmetrical, working interferometer possessing identical high-reflection coatings can be written ($F=\pi$; $r_1^2=r_2^2=R$; $t_1^2=t_2^2=T$)

$$I_{\mathrm{RX}}=R+\left\{\frac{T^2R+2TR^2-2TR\cos\Delta}{1+R^2-2R\cos\Delta}\right\}.$$

The maxima occur at $\Delta=(2n+1)\pi$ and the minima at $\Delta=2n\pi$.

Minima. The minimum intensity in the reflected pattern $[I_{\mathrm{Rx}}]_{\mathrm{min}}$ is, therefore,

$$[I_{\mathrm{Rx}}]_{\mathrm{min}} = R + \frac{RT}{(1-R)^2}\{T+2R-2\}.$$

If there is no absorption in the silver films we can write $R+T=1$, and the minimum reflected intensity in the case of no absorption, $[I_{\mathrm{Rx}}]^{0}_{\mathrm{min}}$, becomes zero. If there is absorption defined by A, such that $R+T+A=1$, the minimum reflected intensity $[I_{\mathrm{Rx}}]^{\mathrm{A}}_{\mathrm{min}}$ becomes $A^2R/(1-R)^2$. Thus

$$[I_{\mathrm{Rx}}]^{\mathrm{A}}_{\mathrm{min}} = [I_{\mathrm{Rx}}]^{0}_{\mathrm{min}} + \frac{R}{(1-R)^2}A^2.$$

Therefore the minimum intensity in the reflected pattern is raised from its value zero in the case of no absorption to the value $A^2R/(1-R)^2$ when there is absorption defined by $R+T+A=1$.

Maxima. The maximum intensity in the reflected pattern is

$$[I_{\mathrm{Rx}}]_{\mathrm{max}} = R + \frac{RT}{(1+R)^2}\{T+2R+2\}.$$

If there is no absorption we can write $R+T=1$, and the maximum reflected intensity becomes

$$[I_{\mathrm{Rx}}]^{0}_{\mathrm{max}} = R + \frac{R}{(1+R)^2}\{3-2R-R^2\}.$$

If there is absorption defined as before, the maximum reflected intensity for absorption A, $[I_{\mathrm{Rx}}]^{\mathrm{A}}_{\mathrm{max}}$, is

$$[I_{\mathrm{Rx}}]^{\mathrm{A}}_{\mathrm{max}} = R + \frac{R}{(1+R)^2}\{(3-2R-R^2)+(A^2-4A)\}.$$

Since typical values of A at the higher reflectivities employed in multiple beam interferometry are of the order of 0·05, $A^2 \ll 4A$. Therefore

$$[I_{\mathrm{Rx}}]^{\mathrm{A}}_{\mathrm{max}} = R + \frac{R}{(1+R)^2}\{(3-2R-R^2)-4A\}.$$

Thus

$$[I_{\mathrm{Rx}}]^{\mathrm{A}}_{\mathrm{max}} = [I_{\mathrm{Rx}}]^{0}_{\mathrm{max}} - \frac{R}{(1+R)^2}4A.$$

Therefore when absorption A is present, defined by $R+T+A=1$, the maximum reflected intensity is decreased by an amount $4AR/(1+R)^2$.

The usefulness of transmitted fringes is usually described by reference to their "half-width" or the reciprocal of the half-width, which is a measure of sharpness, and to the quantity $\{I_{\mathrm{max}} - I_{\mathrm{min}}\}$, which is a measure of their visibility. Similar indices can be used to describe the reflected system at high reflection coefficients. The following Table (Tolansky 1946) gives typical values of R, T and A for silver films of quite low absorption as used in topographical studies (green light).

R (%)	60	70	75	80	85	90	94
T (%)	35·5	27	22	16·5	10·5	4·5	0·7
A (%)	4·5	3	3	3·5	4·5	5·5	5·3

Figure 7 shows the values of $[I_{\mathrm{Rx}}]_{\mathrm{min}}$ and $[I_{\mathrm{Rx}}]_{\mathrm{max}}$ for these good practical values of R, T and A in the curves marked 1, while Figures 8 and 9 show respectively the "sharpness" and "contrast" of the reflected and transmitted fringe systems from a symmetrical interferometer using these silverings.

For comparison with reflected fringes it may be noted that the transmitted fringes from silverings of reflectivity 93% have a half-width of 1/40 but a maximum intensity of only 5% of the incident intensity.

Absorption. From the sharpness and contrast curves of the transmitted and reflected systems the superiority of the reflected fringes for precision measurements is apparent. The preparation of the silver films for use with these systems, however, requires extra care, as the following considerations will show. In Figure 7 are drawn two families of curves: the upper set shows the value of the maximum reflected intensity $[I_{Rx}]_{max}$ for different values of the absorption A; the lower set shows the value of the minimum reflected intensity $[I_{Rx}]_{min}$ for the same values of the absorption. In each family the curve marked 0 shows the values of the quantities when A is very small, $\sim 0{\cdot}01$, the curve marked 1 shows

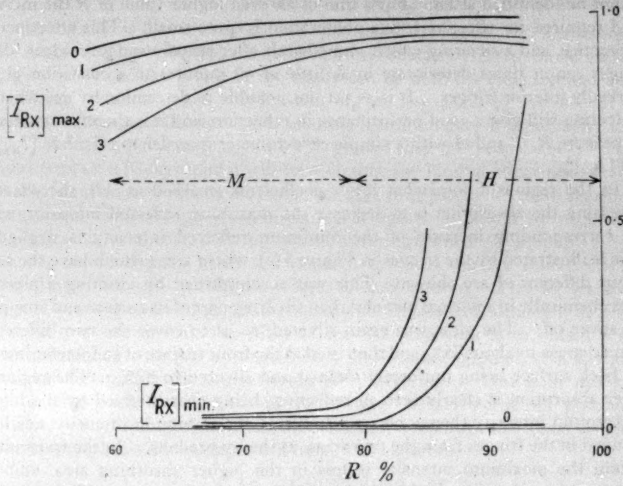


Figure 7.

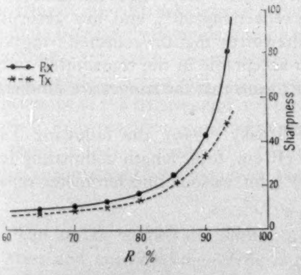


Figure 8.

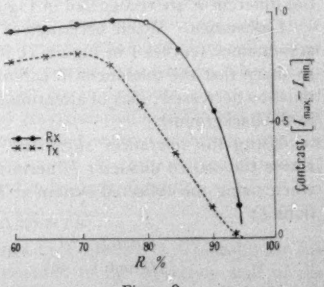


Figure 9.

the values of the quantities when A has the values shown in the Table above, and the curves marked 2 and 3 show the quantities when the absorption has values respectively two and three times these good practical values. In the region of high coefficients (marked as H), it can be seen that a slight increase in the value of A for a film of reflectivity 90% changes conditions critically from sharp fringes of good contrast to conditions in which the contrast is so poor that the fringes cannot be identified at all. For a film of an even higher value of R the increase in A required for effective fringe obliteration is quite small. This effect occurs in practice, and a silvering which immediately after preparation gives clear, sharp fringes can at times deteriorate in as little as 60 minutes to a condition giving markedly inferior fringes. It is as yet not possible to determine by eye whether a silvering will give a good performance in reflection, and it is therefore necessary to measure R, T and A with a simple reflectometer in order to calculate $[I_{\mathrm{Rx}}]^{\mathrm{A}}_{\mathrm{max}}$ and $[I_{\mathrm{Rx}}]^{\mathrm{A}}_{\mathrm{min}}$.

In the region of somewhat lower coefficients (marked as M), the effect of increasing the absorption is to decrease the maximum reflected intensity, while the corresponding increase of the minimum reflected intensity is negligible. This is illustrated by the fringes in Figure 5 (*c*), where areas which have the same R but different A are shown. This was accomplished by cleaning a piece of glass chemically in a normal manner, then silvering part of its surface and stripping the silver off. The glass was again silvered, so as to cover the two differently cleaned areas to about 75% and then used as the front surface of an interferometer, the back surface being uniformly cleaned and silvered to 82%. The region of lower absorption is clearly seen in reflection, being characterized by the higher background intensity (hence better visibility); the minimum intensity was little changed in the fringes from the two areas, as theory predicts. In the transmitted system the maximum intensity is less in the higher absorbing area, and the fringes are made to look sharper by relative under-exposure. The stripping process is thus seen to be an effective cleaning process. Areas of the same A but different R are recognized in Figure 5 (*c*) by the "step" between them.

Collimation. When silverings of high reflectivity 90% and low absorption are obtained (curves 1 in Figure 7), it is noteworthy that the reflected fringes are so sharp that the tolerances in collimation acceptable in the transmitted system must be decreased; lack of attention to this means that the fringes are submerged in the background.

Using the tolerances suggested by Tolansky (1946), the following Table shows the source diameter D permissible (10 cm. focal length collimating lens) when using the reflected system at $R = 90\%$ for various interferometer separations t:

t (mm.)	1	0·1	0·01	0·001
D (mm.)	0·1	0·35	1·00	3·000

§ 8. UNSILVERED BACK SURFACES

Although it is desirable to have high reflecting coefficients within the limits discussed when using the reflected system in surface studies, cases arise when the back surface of the interferometer, which is the surface under examination, will not be coated with a highly reflecting film since the several techniques by which this is achieved may influence undesirably the surface structure. In this case, the natural reflectivity of the surface under examination will be used and will, in general, be considerably lower than that of silver. The effect of having R_2 small is to cause the shape factor Θ_Δ to have less rapid variations about $\Delta = 2n\pi$, and to decrease the value of the magnitude terms upon which it operates, in comparison with the background term R_1, as the expression for the reflected intensity from an asymmetrical interferometer (R_1, R_2) shows :

$$I_{\mathrm{RX}} = \left\{ R_1 + \frac{T_1^2 R_2 + 2T_1\sqrt{(R_1R_2)}\cos(\Delta - F) - 2T_1R_1R_2\cos F}{1 + R_1R_2 - 2\sqrt{(R_1R_2)}\cos\Delta} \right\}.$$

The effect of this is loss of contrast and sharpness. However, by decreasing R_1 to the range of values 60–70% the asymmetrical reflection fringes can be usefully employed since the sharp side of the fringe can be used for measurement while the contrast remains good. Even with R_2 as low as 25% the use of the asymmetrical fringes will allow quite accurate measurements of topography to be made. Since the reflected fringes with R_1 greater than 60% consist of some kind of minimum on a background it will be seen that any kind of effect which causes the background to encroach on the fringe will give an apparent increased sharpness. For instance, in the range of asymmetrical fringes it is found by experiment and confirmed by theory that for typical silverings when R_1 is about 55% (a value dependent slightly upon the value of the back surface reflectivity R_2), the minimum intensity within the fringes is close to zero. This allows a high degree of over-exposure when photographing the fringes and, although the fringes are actually asymmetrical, the effect of the over-exposure is to obtain a final image resembling a simple minimum on a uniform background, of a sharpness somewhat resembling the fringes from 80% silverings. This is of use when dealing with back surfaces of low reflectivity. Other particular procedures are of use when dealing with certain surfaces, and a report of the actual investigations on several metallic surfaces will be published later.

§ 9. CONCLUSION

From the point of view of topographical studies, the characteristics of the reflected system may be summed up briefly. For reflectivities greater than 90% the fringes are sharper than the corresponding transmitted fringes but their visibility is critically dependent upon the absorption in the silver film; collimation conditions are strict. Complications of the simple interference relationship due to focusing defect and large wedge angles are immediately obvious in the appearance of the fringes, and one can avoid the errors that occur in transmission work if high precision measurements be attempted without sufficient attention to these factors. Finally, the results given apply to the reflected fringes of equal chromatic order which obey the same fundamental formula.

ACKNOWLEDGMENTS

The author wishes to express his thanks to Professor S. Tolansky for his interest and encouragement; he is also indebted to the workshop staff of the Physics Department, Royal Holloway College, for much help and to the Ministry of Education for a Forces Education Grant, which enabled the work to be carried out.

REFERENCES

Brossel, J., 1947, *Proc. Phys. Soc.*, **59,** 224.

Hamy, M., 1906, *J. Phys. Radium*, **5,** 789.

Rouard, P., 1937, *Ann. Phys., Paris*, **7,** 291.

Tolansky, S., 1945, *Proc. Roy. Soc.* A, **184,** 51; 1946, *Proc. Phys. Soc.*, **58,** 684.

Tolansky, S., and Wilcock, W. L., 1947, *Proc. Roy. Soc.* A, **191,** 182.

Optical Transmission of a Colloidal Solution in a Magnetic Field

By F. D. STOTT
Electrical and Clarendon Laboratories, Oxford University

Communicated by A. von Engel; *MS. received* 28*th July* 1948, *and in amended form* 15*th February* 1949

ABSTRACT. The change in optical transmission of a solution of colloidal graphite in a magnetic field has been measured as a function of a field strength and of concentration, and the results obtained compared with those to be expected from theory.

It is found that the experimental results agree closely with the hypothesis that the change in transmission is due to the change in projected area of the plate-like particles when these are aligned by the magnetic field.

By comparison of theoretical and experimental curves the particle volume is found. The rate of decay, after removal of the field, of the change in transmission produced by the field has also been measured.

An approximate theory is given to connect this rate with the diameter of the particles, and from this the size of the particles has been calculated. The dimensions so found agree satisfactorily with evidence from other sources.

---

§ 1. INTRODUCTION

THE change in optical transmission of a medium consisting of a suspension of small particles in a liquid produced by application of a magnetic field in the same direction as the light beam was first observed qualitatively by W. R. Groves in 1845 (McKeehan 1940). He used for his experiment a suspension of ferromagnetic particles in water. This effect has recently been the subject of several papers (Heaps 1940, Elmore 1941, Mueller and Shamos 1942), but the theory remains incomplete owing to the doubt which exists about the magnetic properties of these very small ferromagnetic particles; nor has a complete experimental investigation been carried out.

A similar effect has been observed by Stuart (1938) in a solution of colloidal graphite. In the course of experiments which he carried out to find a method of detecting and estimating small quantities of graphite in oil, Stuart found that the optical transmission of such a solution changed when a magnetic field was applied, and he made approximate measurements at a fixed field strength for various concentrations.

As graphite has known magnetic properties, being diamagnetic and anisotropic, it is possible to develop a satisfactory theory of this effect. The object of this paper is to describe experimental results obtained, and to show what information about the size and shape of the graphite crystals can be obtained by comparing the theoretical and experimental results.

§ 2. THEORY

(i) *The Change in Optical Transmission as a Function of Field Strength and Concentration*

A solution of colloidal graphite in water forms a true colloid system, in that the particles remain in permanent suspension and exhibit Brownian motion. As the particles are disc-shaped, the absorption due to any one particle will depend on its orientation with respect to the light beam, and may be assumed proportional to its projected area; the diffraction effects which will also enter into the total absorption may be neglected here, as they may be assumed independent of orientation, and we are concerned here only with that part of the absorption which is dependent on orientation.

In order to find the relation between absorption, or transmission, and the applied magnetic field strength, it is first necessary to find the distribution of orientation as a function of field strength. This is determined by two factors—the torque exerted on the particle by the magnetic field, and the disturbing effect of the Brownian motion, which acts to prevent perfect alignment, tending always to re-establish random orientation.

Distribution of orientation as a function of field strength. The orientation of a particle may be defined by two coordinates giving the position of the end of the normal to the basal plane on the surface of a unit sphere whose polar axis is parallel to the applied field. The coordinates used here are: θ, co-latitude, and ϕ, longitude. The torque acting on a particle of volume V whose susceptibilities are $-\chi_1$, $-\chi_2$ normal to the basal plane and in the basal plane respectively is given by (see Figure 1)

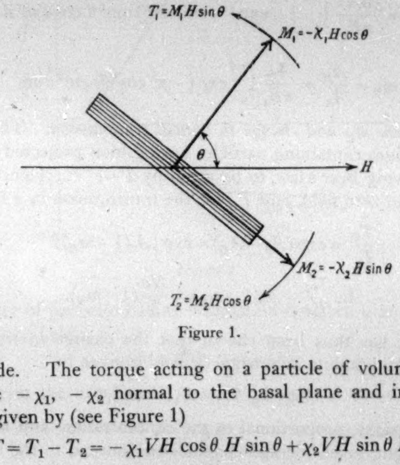


Figure 1.

$$T = T_1 - T_2 = -\chi_1 VH \cos\theta \, H \sin\theta + \chi_2 VH \sin\theta \, H \cos\theta$$
$$= (\chi_2 - \chi_1) H^2 V \sin\theta \cos\theta,$$

where H is the magnetic field strength. The potential energy (rotational) is therefore given by

$$U = \int T \, d\theta = \tfrac{1}{2}(\chi_1 - \chi_2) H^2 V \cos^2\theta.$$

From Boltzmann's equation, the number of particles having orientation between θ and $\theta + \delta\theta$, and ϕ and $\phi + \delta\phi$ will therefore be

$$\delta N = C \exp\left[-\frac{(\chi_1 - \chi_2)}{2\boldsymbol{k}T} H^2 V \cos^2\theta \right] \sin\theta \, \delta\theta \, \delta\phi,$$

which, after integrating to find C, gives

$$\frac{\delta N}{N} = \frac{2\alpha \exp(-\alpha^2 \cos^2\theta) \sin\theta \, \delta\theta \, \delta\phi}{\pi E_\alpha}, \qquad \ldots\ldots(1)$$

where $\alpha^2 = \frac{1}{2\boldsymbol{k}T}(\chi_1 - \chi_2) H^2 V$, and $E_\alpha = \int_0^\alpha e^{-x^2} dx$.

Light parallel to field. For a beam of light parallel to the applied field, the projected area of a particle of area a is $a\cos\theta$, assuming the thickness negligible compared with the diameter. The total projected area of the N particles per unit volume is therefore

$$A_H = \frac{2aN\alpha}{\pi E_\alpha}\int_0^{\pi/2}\int_0^{\pi/2} \exp(-\alpha^2\cos^2\theta)\cos\theta\sin\theta\,d\theta\,d\phi. \qquad \ldots\ldots(2)$$

Integrating this with respect to ϕ and θ, we have $A_H = Na\{1-\exp(-\alpha^2)\}/2\alpha E_\alpha$. For zero field strength, $\alpha \to 0$, and, therefore, $A_0 = Na/2$, so that

$$m_H = \frac{A_H}{A_0} = \frac{1-\exp(-\alpha^2)}{\alpha E_\alpha}. \qquad \ldots\ldots(3)$$

Light normal to field. When the light beam passed at right angles to the applied field, the projected area of a particle of area a is $a\sin\theta\cos\phi$.

From (1), therefore,

$$A'_H = \frac{2aN\alpha}{\pi E_\alpha}\int_0^{\pi/2}\int_0^{\pi/2} \exp(-\alpha^2\cos^2\theta)\sin^2\theta\cos\phi\,d\theta\,d\phi, \qquad \ldots\ldots(4)$$

which gives

$$m'_H = \frac{A'_H}{A_0} = \frac{4\alpha}{\pi E_\alpha}\int_0^{\pi/2} \exp(-\alpha^2\cos^2\theta)\sin^2\theta\,d\theta. \qquad \ldots\ldots(5)$$

Relation between m_H and change in optical transmission. The optical transmission of a medium containing particles having total projected area A is easily shown, as in deriving Beer's law, to be given by $T = e^{-A}$. Hence, writing T_0 for the transmission in zero field, and T_H for the transmission in a field H, we have

$$\frac{T_H}{T_0} = \exp(A_0 - A_H) = \exp\{A_0(1-m_H)\}.$$

Hence

$$\ln\frac{T_H}{T_0} = A_0(1-m_H) = \frac{Na}{2}(1-m_H).$$

Summary. It has thus been shown that the change in transmission of a solution of colloidal graphite in a magnetic field is given by

$$\log(T_H/T_0) = K(1-m_H), \qquad \ldots\ldots(6)$$

where K is a constant, proportional to the concentration, and m_H is a function of the field strength, given for light parallel and normal to the applied field by equations (3) and (5) respectively.

Graphs of $(1-m_H)$ against the parameter α are shown in Figure 2.

Equation (6) also shows that for a constant field strength $\log(T_H/T_0)$ is proportional to concentration.

(ii) *Theory of Relaxation Time*

If the particles are aligned by a magnetic field, and this field is then removed suddenly, the Brownian motion of the particles will re-establish random orientation. The time required for the process will be a function of the particle radius. Thus the change in optical transmission produced by the field will decay until the transmission has returned to its normal zero-field value.

Debye (1929) has given a formula for the relaxation time for polar molecules; this formula is not, however, applicable to this case, for two reasons: first, the dependence of the observed effect on angle of orientation is different, and second, Debye's formula applies only to small degrees of alignment, i.e. cases in which the curve of distribution versus angle of orientation is broad. An approximate formula for high degrees of orientation is therefore derived here, following Einstein's method for the case of diffusion from a point (Einstein 1906).

At the moment of removing the field, the particles are aligned with their basal planes parallel to the field, and the coordinate θ defined above has the value $\frac{1}{2}\pi$ for this position. It is more convenient now to take a new coordinate $\psi = \frac{1}{2}\pi - \theta$.

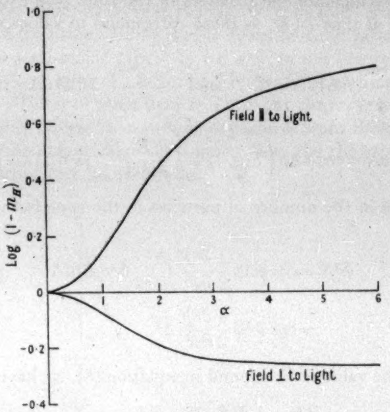


Figure 2.

If the number of particles having orientation between ψ and $\psi + \delta\psi$ at time t after removal of the field is

$$\delta N = F(\psi, t) \cos\psi \, \delta\psi, \qquad \ldots\ldots(7)$$

then the change in the number of particles between ψ and $\psi + \delta\psi$ in a further short time δt is

$$\delta^2 N = \delta t \, \frac{\partial F}{\partial t} \cos\psi \, \delta\psi. \qquad \ldots\ldots(8)$$

If the number of particles suffering a change of orientation of value between Δ and $\Delta + \delta\Delta$ in time δt is

$$\delta N = N f(\Delta) \delta\Delta,$$

then the number of particles entering the zone between ψ and $\psi + \delta\psi$ from the zone between $\psi + \Delta$ and $\psi + \Delta + \delta\Delta$ in the time δt is given by

$$\begin{aligned} \delta^2 N &= F(\psi + \Delta, t) \cos(\psi + \Delta) \delta\Delta f(\Delta) \, \delta\psi \\ &= \cos\psi \, F(\psi + \Delta, t) f(\Delta) \, \delta\Delta \quad \text{approx.,} \qquad \ldots\ldots(9) \end{aligned}$$

$\cos(\psi + \Delta)$ being replaced by $\cos\psi$ in the approximation, since, provided δt is small, only small values of Δ need be considered. This approximation is equivalent to regrading each elementary zone of the spherical surface as a section of a cylinder.

Expanding $F(\psi+\Delta, t)$ and integrating over all values of Δ, we obtain for the total number of particles in the zone between ψ and $\psi+\delta\psi$ at time $t+\delta t$

$$\delta N = \cos\psi\,\delta\psi\left[F(\psi, t)\int_{-\pi/2}^{\pi/2} f(\Delta)\,d\Delta + \frac{\partial F}{\partial\psi}\int_{-\pi/2}^{\pi/2}\Delta f(\Delta)\,d\Delta + \frac{1}{2}\frac{\partial^2 F}{\partial\psi^2}\int_{-\pi/2}^{\pi/2}\Delta^2 f(\Delta)\,d\Delta + \ldots\right].$$

This automatically includes the particles in the zone ψ, $\psi+\delta\psi$ at time t which still remain there at time $t+\delta t$, as these correspond to values of Δ less than $\delta\psi$. Since

$$\int_{-\pi/2}^{\pi/2} f(\Delta)\,d\Delta = 1, \qquad \text{and} \qquad \int_{-\pi/2}^{\pi/2}\Delta f(\Delta)\,d\Delta = 0,$$

this reduces to

$$\delta N = \cos\psi\,\delta\psi\left(F(\psi, t) + \frac{1}{2}\frac{\partial^2 F}{\partial\psi^2}\int_{-\pi/2}^{\pi/2}\Delta^2 f(\Delta)\,d\Delta\right), \qquad \ldots\ldots(10)$$

so that the change in the number of particles in the zone between ψ and $\psi+\delta\psi$ in time δt is

$$\begin{aligned}\delta^2 N &= \cos\psi\,\delta\psi\,\frac{1}{2}\frac{\partial^2 F}{\partial\psi^2}\int_{-\pi/2}^{\pi/2}\Delta^2 f(\Delta)\,d\Delta \\ &= \cos\psi\,\delta\psi\,\frac{1}{2}\frac{\partial^2 F}{\partial\psi^2}\,\overline{\Delta^2}.\end{aligned}$$

Equating this to the value of $\delta^2 N$ found in equation (8), we have

$$\frac{\partial F}{\partial t} = \frac{1}{2}\frac{\partial^2 F}{\partial\psi^2}\frac{\overline{\Delta^2}}{\delta t} = \frac{\partial^2 F}{\partial\psi^2}\,\boldsymbol{k}TB, \qquad \ldots\ldots(11)$$

since from Einstein's equation for the rotational Brownian movement, $\frac{1}{2}(\Delta^2/\delta t) = \boldsymbol{k}TB$, where $\boldsymbol{k}$ is Boltzmann's constant, T the absolute temperature, and B the rotational mobility (i.e. angular velocity per unit torque).

The solution to equation (11) is

$$F(\psi, t) = Ct^{-\frac{1}{2}}\exp(-\psi^2/\beta t), \qquad \text{where} \quad \beta = 4\boldsymbol{k}TB.$$

Hence $\qquad \delta N = Ct^{-\frac{1}{2}}\exp(-\psi^2/\beta t)\cos\psi\,\delta\psi$

and

$$\frac{\delta N}{N} = \frac{\exp(-\psi^2/\beta t)\cos\psi\,\delta\psi}{\int_0^{\pi/2}\exp(-\psi^2/\beta t)\cos\psi\,d\psi},$$

from which the projected area of the particles at time t is given by

$$A_t = \frac{Na\int_0^{\pi/2}\exp(-\psi^2/\beta t)\cos\psi\sin\psi\,d\psi}{\int_0^{\pi/2}\exp(-\psi^2/\beta t)\cos\psi\,d\psi} = \frac{Na\mathscr{I}_1}{\mathscr{I}_2}.$$

For large values of t, $\mathscr{I}_1/\mathscr{I}_2 \to \frac{1}{2}$, so that $A_t \to \frac{1}{2}Na$, which is the normal value in zero field; calling this A_∞, we have $m_t = A_t/A_\infty = 2\mathscr{I}_1/\mathscr{I}_2$. The quantity m_t is connected with the transmission by a similar relation to that given for m_H in §2 (i) above, i.e. $\log(T_t/T_\infty) = K(1-m_t)$.

In Figure 3, $\log(1-m_t)$ is plotted as a function of βt; from $\beta t = 0{\cdot}2$ onward it approximates to a straight line, showing that $(1-m_t)$ decreases exponentially, the mean time constant being given by $\beta\tau = 0{\cdot}8$, or $\tau = 0{\cdot}8/\beta$ approx. Hence it is to be expected that $\log(T_t/T_\infty)$ will decrease exponentially after the field is removed. Before this relation can be used to determine the particle radius R, it is necessary to fix the value of B, in which R is contained; for spherical particles $B = 1/8\pi\eta R^3$. Perrin (1934) has compared the values of B for ellipsoidal particles with those for spherical particles of the same volume. For an ellipsoid of revolution whose major and minor axes are in the ratio 5 : 1, he finds B about 2·5 greater than for a sphere of the same volume, hence about half the value for a sphere of radius equal to the semi-major axis. As we wish to use the relaxation time to find R, and $R \propto B^{-\frac{1}{3}}$, an error of a factor of 2 in the value assumed for B will not affect the result by a factor of more than $2^{\frac{1}{3}}$, i.e. 1·26; the error in using the value $1/8\pi\eta R^3$ will probably be less than this, as departures from Stokesian conditions will tend to make the actual value of B larger than the theory indicates. The value $1/8\pi\eta R^3$ will therefore be used here.

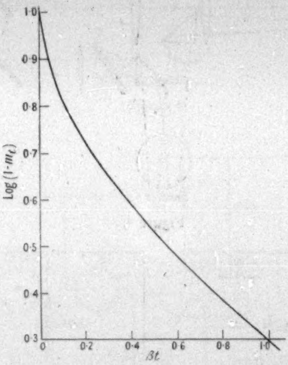


Figure 3.

The time constant τ will then become $\tau = 0{\cdot}8 \times 2\pi\eta R^3/\boldsymbol{k}T$, which gives

$$R = \left(\frac{\boldsymbol{k}T\tau}{1{\cdot}6\pi\eta}\right)^{\frac{1}{3}}.$$

§ 3. EXPERIMENTAL METHODS AND RESULTS

(i) *Experimental Methods*

When the applied magnetic field is small, and the concentration low, the change in transmission may be only a small fraction of the transmission in zero field. It is necessary therefore to use a method of measuring the change in which the effect due to the magnetic field will not be masked by fluctuations in the intensity of the light source or in the sensitivity of the photocell. This is achieved by using a comparison method, in which the intensity of the light beam passed through a cell containing the specimen of colloidal graphite is compared with the intensity of another beam from the same source, the second beam being adjusted to be equal to that through the graphite in zero field. Two different methods of comparison have been used in these experiments; the first uses a single photocell to compare the two beams, which are switched on to it alternately by a rotating glass disc, half of which is silvered and half clear. If the two beams are unequal in intensity, an alternating voltage of square waveform appears across the photocell load resistance, of amplitude proportional to the difference of intensity between the two beams. This is amplified and applied to the Y plates

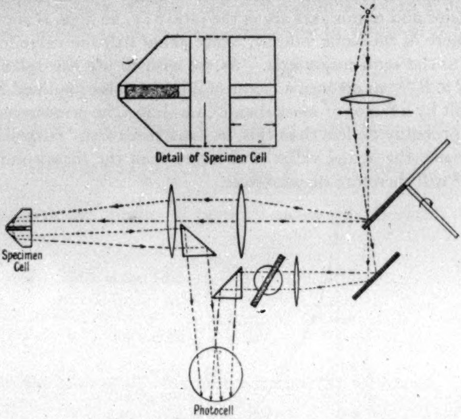

Detail of Specimen Cell
Specimen Cell
Photocell

Figure 4.

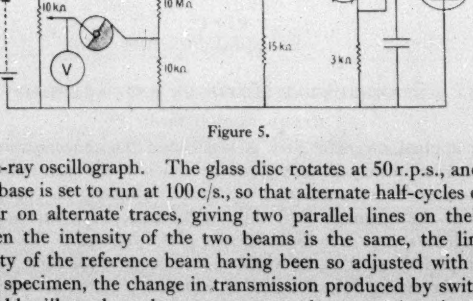

+100 v Stabilized
CV 242
470 kΩ
10 MΩ
10 kΩ
10 kΩ
V
6J5
SP41
100 kΩ
15 kΩ
3 kΩ
+250 v Stabilized

Figure 5.

of a cathode-ray oscillograph. The glass disc rotates at 50 r.p.s., and the oscillograph time-base is set to run at 100 c/s., so that alternate half-cycles of the square wave appear on alternate traces, giving two parallel lines on the cathode-ray tube. When the intensity of the two beams is the same, the lines coincide. The intensity of the reference beam having been so adjusted with no magnetic field on the specimen, the change in transmission produced by switching on the magnetic field will produce the square-wave voltage, causing the two traces to separate. The amplitude of the voltage produced at the photocell is then measured by applying another square-wave voltage, in the opposite phase, to the lower end of the photocell load resistance, the amplitude of this voltage being adjusted to make the traces coincide again. This voltage is derived from a D.C. supply via a commutator driven on the same shaft as the glass disc, and it is the D.C. voltage which is actually measured. This method gives greater accuracy than can be obtained by attempting to measure the voltage of the original square wave directly, and also ensures that the voltage across the photocell is maintained constant, although the two beams are of different intensity.

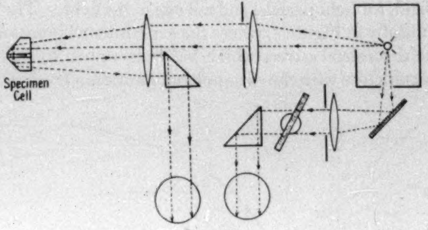


Figure 6.

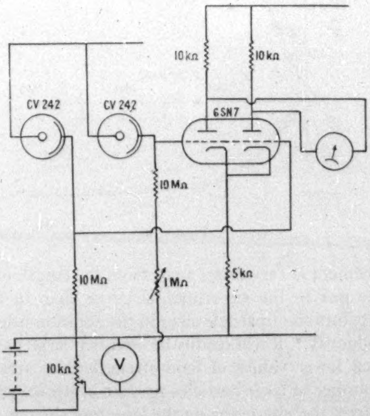


Figure 7.

The second method of comparison uses two photocells, one for each beam. The outputs from the two cells are amplified by the two halves of a double triode, between the anodes of which is connected a milliammeter. The equality of the two beams is indicated by zero current on the milliammeter. The change in ntensity of the beam through the colloid specimen is measured by finding the voltage which has to be applied to the lower end of one of the photocell load resistances to restore the balance. Thus in this method, as in the previous one, the voltage across the photocells is kept constant.

The optical and electrical arrangements of the first method are shown diagrammatically in Figures 4 and 5, and of the second method in Figures 6 and 7.

(ii) *Experimental Results*

Of the two methods described above, the first is rather more accurate, due to the absence of drift and slow fluctuations, which is the result of using an A.C. amplifier. This was therefore used to measure the change in optical transmission as a function of the applied field, for light parallel and normal to the field. The results obtained are shown graphically in Figure 8, where the experimental points are shown superimposed on the theoretical curve, which has been drawn to such a scale that its saturation level coincides with the extrapolated saturation level of the experimental points. The agreement is very close, apart from the length of the toe at low field strength. This is less in the experimental curve than in the theoretical one, because of the distribution of particle sizes in the solution used; as the parameter α contains the product H^2V it will readily be seen that large particles will approach saturation at much lower values of field strength than small ones; hence the effect of a small number of large particles present in the solution is to steepen the first part of the curve, thereby reducing the length of the toe.

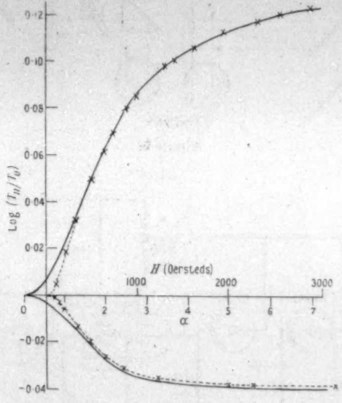


Figure 8.

Effect of varying the concentration. The second experimental method described above was used to study the effect of varying the concentration. Although this method is not quite so accurate as the first one, it is more readily adapted for measuring the transmission factor T_0; this was not required in the first experiment, as relative values of the change produced by the field were all that was required. In these experiments, however, accurate values of T_0 are required since, the concentration being very small (of order 10^{-4} by weight), more accurate results may be obtained by measuring T_0 and assuming the validity of Beer's law than by measuring the concentration directly.

The results are shown in Figures 9 and 10. In Figure 9, $\log(T_H/T_0)$ is plotted against H for each concentration; it will be seen that all the curves are similar, as is to be expected. In Figure 10, $\log(T_H/T_0)$ is plotted against $\log T_0$ ($\propto$ concentration) for various field strengths. There is some falling off from the direct proportionality predicted by the theory at higher concentrations; this may be due to increasing importance of diffraction effects.

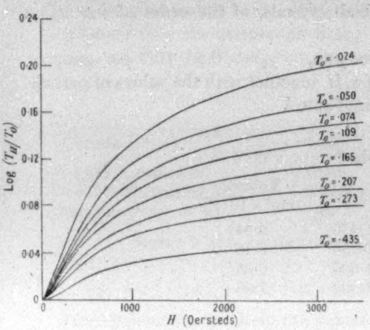


Figure 9.

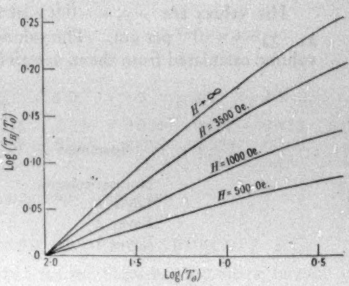


Figure 10.

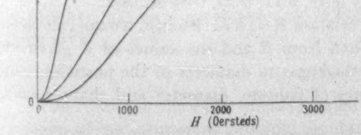


Figure 11.

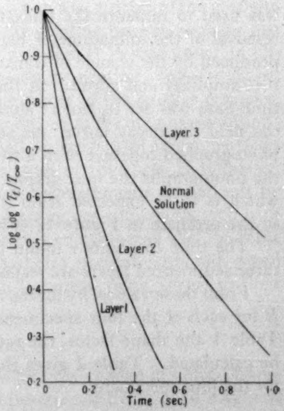


Figure 12.

Different particle sizes. A solution of colloidal graphite of concentration of order 10^{-3} by weight was allowed to sediment in a tall vessel, and then separated into a number of layers. In Figure 11, $\log(T_H/T_0)$ is plotted against H for three of these layers, Layer 1 being the top, Layer 2 a central one, and Layer 3 the bottom. It is noticeable, particularly in the cases of Layers 1 and 2, that the greater homogeneity of particle size in these specimens as compared with the normal solution has lengthened the toe of the curve, making it similar to that predicted by the theory.

Calculation of particle volume from these results. The parameter α defined in §2 (i) is $[(\chi_1-\chi_2)/2\mathbf{k}T]^{\frac{1}{2}}HV^{\frac{1}{2}}$, which is proportional to H. By comparing the experimental curves in Figures 8 and 11 with the theoretical curve for $\log(T_H/T_0)$ we can obtain a value of the factor $\alpha/H=[(\chi_1-\chi_2)/2\mathbf{k}T]^{\frac{1}{2}}V^{\frac{1}{2}}$. The values of $\mathbf{k}$, Boltzmann's constant, and T, the absolute temperature, are known, and Ganguli (1936) gives values for χ_1 and χ_2 for small crystals, of the order of size to be expected here.

His values are $-\chi_1=-0{\cdot}9\times10^{-6}$ per gm., $-\chi_2=-0{\cdot}5\times10^{-6}$ per gm., or $\chi_1-\chi_2=4\times10^{-7}$ per gm. The values of α/H, together with the values of particle volume calculated from them, are given in Table 1.

Table 1

Specimen	α/H	Volume ($\text{cm}^3\times10^{-12}$)
Normal solution	1/450	0·44
Layer 1	1/750	0·16
Layer 2	1/600	0·25
Layer 3	1/300	1·00

(iii) *Relaxation Time Measurements*

Method of measurement. The second form of the apparatus described above was used to measure the relaxation time. The only change required was the removal of the milliammeter between the triode anodes; the voltage difference produced by the change in transmission due to the field was then amplified by a D.C. amplifier and applied to the Y plates of a cathode-ray oscillograph. The time-base was set to run at about 3–5 cycles per second, and on switching off the field the decay curve was seen spread out over several traces. This was photographed together with a timing trace, and from measurements taken from the photographs the four graphs in Figure 12 have been plotted. The quantity which is to be expected from the theory to decay exponentially is $\log(T_t/T_\infty)$, so the ordinate in Figure 12 is $\log\log(T_t/T_\infty)$, plotted against time in seconds.

The time constants τ found from Figure 12 for the normal solution and the three sedimented layers are respectively 0·59, 0·17, 0·29, 0·89 second.

From these values, by means of the relation $R=(\mathbf{k}T\tau/1{\cdot}6\pi\eta)^{\frac{1}{3}}$, we can calculate R for each of the four specimens. Then from R and the values of V given in Table 1 the shape factor, the ratio of thickness to diameter of the particles, can be calculated. Table 2 gives the values of volume, diameter and shape factor for the four specimens.

Table 2

Specimen	Volume ($\text{cm}^3\times10^{-12}$)	Diameter (μ)	Shape factor
Normal solution	0·44	1·56	1/6·9
Layer 1	0·16	1·04	1/5·45
Layer 2	0·25	1·24	1/5·95
Layer 3	1·00	1·80	1/4·6

§ 4. A.C. MAGNETIC FIELD

Since the torque acting on the particles is proportional to H^2, it is to be expected that an alternating field will have a similar effect to a steady field. This was found to be the case, and in Table 3 comparison is made between steady and alternating fields, the change in transmission here being expressed as a percentage. In the case of the alternating current, the current given is the peak value, not r.m.s It will be seen that at fairly large field strengths the change in transmission produced by the A.C. field is nearly 90% of that produced by a steady field equal to the peak value of the alternating field. This is due to the fact that the relaxation time is much longer than the orientation time; the latter decreases with increasing field strength, whereas the former remains constant.

Table 3

Current (amp.) D.C. / A.C. peak	0·25	0·5	1·0	1·5	2·0	2·5
(*a*) $\delta T/T$, steady field	0·066	0·145	0·256	0·334	0·381	0·413
(*b*) $\delta T/T$, alternating field	0·0326	0·088	0·166	0·255	0·323	0·360
(*b*) as percentage of (*a*)	49	61	65	76	84	87

NOTE: 1 amp. corresponds to approximately 1,000 gauss.

§ 5. EXAMINATION OF PARTICLES UNDER MICROSCOPE

An independent check on the values found for the diameter and shape factor of the particles in the experiments described above was obtained by examination under a microscope. The colloidal graphite was mixed with gelatine and allowed to set in a magnetic field, being slowly rotated about an axis normal to the particles with their basal planes all parallel; thus when the gelatine has set, sections can be cut normal to the basal planes. When such a section is viewed under an oil-immersion microscope, the particles are seen edge-on, presenting a rod-like shape. It is not of course possible to measure the shape factor of the smaller particles, of order 1 μ or less in diameter, as these, seen edge-on, are at the limit of resolution; approximate measurements on a few of the small proportion of larger particles (2–5 μ) present gave values of about 1/5 for the ratio of thickness to diameter.

The average particle diameter is known, from measurements carried out by Acheson Colloids Limited, to be about 1·1 μ, which agrees sufficiently well with the results obtained in these experiments, when it is remembered that, in these optical absorption measurements, the larger particles present, although small in number, have a disproportionate influence on the mean value, owing to their large area.

§ 6. CONCLUSION

The results of the experiments carried out to find the change in optical transmission of the solution of colloidal graphite as a function of the applied magnetic field agree well with theory given in § 1. This theory is based on the assumption that the change in optical transmission depends only on the change in projected area which occurs when the particles are aligned by the field. From the agreement found between theory and experiment it appears that this assumption is largely justified.

The value found for the mean particle volume from these experiments is $0{\cdot}44 \times 10^{-12}$ cm^3, and from measurement of the relaxation time the mean diameter is found to be $1{\cdot}56\,\mu$. Both these values may be expected to be larger than the averages by number found from direct microscopic measurement of the particles, as the larger particles, having greater area, will play a predominant part in determining the change in optical transmission produced by alignment; thus the mean value will be displaced in the direction of larger particle size. The value given for the average particle diameter by Acheson Colloids Limited, whose preparation Aquadag 'S' was used for these experiments, is $1{\cdot}1\,\mu$. This agrees sufficiently well with the value $1{\cdot}56\,\mu$ found here, when this displacement is taken into consideration.

The value of about 7 : 1 found for the ratio of diameter to thickness appears reasonable, being of the same order as that usually found for macroscopic crystals of graphite. It is noticeable that this ratio remains approximately constant over the range of particle sizes examined here.

From the results of these experiments it can be concluded that reasonably accurate information about the size and shape of colloid particles can be obtained by studying the change in optical transmission which takes place when the particles are aligned, at any rate in the case of colloidal graphite. It seems likely that similar information could be obtained for other colloids in which alignment effects can be observed, in magnetic or electric fields.

ACKNOWLEDGMENTS

The author's thanks are due to Professor Lord Cherwell, in whose department this work was carried out, and also to Dr. A. von Engel, for much valuable advice and encouragement.

REFERENCES

Debye, P., 1929, *Polar Molecules* (New York : Chemical Catalog Company).
Einstein, A., 1906, *Ann. Phys., Lpz.*, **19**, 371.
Elmore, W. C., 1941, *Phys. Rev.*, **60**, 593.
Ganguli, N., 1936, *Phil. Mag.*, **21**, 355.
Guha, B. C., and Roy, B. P., 1934, *Ind. J. Phys.*, **8**, 345.
Heaps, C. W., 1940, *Phys. Rev.*, **57**, 528.
Heller, W., and Quimfe, G., 1942, *Phys. Rev.*, **62**, 479.
McKeehan, L. W., 1940, *Phys. Rev.*, **57**, 1177.
Mueller, H., and Shamos, M. A., 1942, *Phys. Rev.*, **61**, 631.
Perrin, F., 1934, *J. Phys. Radium*, **5**, 497.
Stuart, A. H., 1938, *Engineering*, **141**, 17.

Electron Flow in Curved Paths under Space-Charge Conditions

By B. MELTZER
Mullard Radio Valve Co. Ltd., Mitcham, Surrey

Communicated by N. W. Robinson: *MS. received 22nd November* 1948

ABSTRACT. A general, synthetic method of obtaining rigorous solutions of steady electron flow subject to space-charge forces is presented. The solutions are not obtained for given boundary conditions, but the boundary conditions are deduced from the solutions. Two examples of such solutions, involving strongly curved two-dimensional electron trajectories, are given; the method is in principle capable of giving the solutions of all possible electron flow patterns in three dimensions except perhaps those involving intercrossing trajectories. It is suggested that the subject offers scope for applied mathematical research at least on the same scale as potential theory.

§1. INTRODUCTION

As far as the writer is aware, no rigorous solutions for electron flow subject to space-charge forces have been published, except for cases involving straight line trajectories. The latter are involved in the classical solutions (Langmuir 1913, Langmuir and Blodgett 1924) for planar, cylindrical and spherical symmetry leading to the Child–Langmuir three-halves power laws, as well as in the more recent application of non-linear mechanics to the problem by Bellustin (1939). It is surprising that other types of motion do not appear to have been thoroughly studied, when one considers the complicated structures of modern electronic devices and how very far from rectilinear their electron trajectories often are. For example, the important question of current division between grid and anode in an oscillator triode can hardly be investigated seriously until the trajectories in the neighbourhood of the grid are known; only then can one take into account such factors as secondary emission from the grid.

This paper presents a synthetic general method for obtaining rigorous solutions of steady electron flow. By "rigorous" we mean that space-charge forces are taken into account; but, as in the original derivation of the Child–Langmuir laws, we shall neglect random (temperature) variations of velocity, assuming that the steady flow is characterized by the existence of a unique velocity vector at every point in the stream. Similarly, we shall assume that particle velocities are so much smaller than the velocity of light that forces due to the magnetic field of the currents may be neglected.

After an exposition of the basic idea of the method, the results for two cases of curvilinear two-dimensional flow are presented, followed by a general discussion of the achievement of these and other solutions.

§2. FUNDAMENTAL EQUATIONS AND METHOD OF SOLUTION

The motion of an electron is governed by the force equation

$$m\dot{\mathbf{v}} = -e\mathbf{E}, \qquad \ldots\ldots(1)$$

where e, m are the magnitudes of the electronic charge and mass, and $\mathbf{v}$, $\mathbf{E}$ are respectively the velocity of the electron and the electric field acting on it.

If we consider the x component of the velocity in a right-handed Cartesian rectangular coordinate system with unit vectors $\mathbf{I}$, $\mathbf{J}$, $\mathbf{K}$ we have

$$d(v_x) = \mathbf{ds}\,\mathrm{grad}\,v_x, \qquad \ldots\ldots(2)$$

where **ds** is a differential displacement in the direction of the motion. Hence

$$\dot{v}_x = \mathbf{v} \text{ grad } v_x. \qquad \text{......(3)}$$

Similarly

$$\dot{v}_y = \mathbf{v} \text{ grad } v_y, \qquad \text{......(4)}$$

and

$$\dot{v}_z = \mathbf{v} \text{ grad } v_z. \qquad \text{......(5)}$$

Substituting these three expressions in equation (1),

$$\mathbf{I}(\mathbf{v} \text{ grad } v_x) + \mathbf{J}(\mathbf{v} \text{ grad } v_y) + \mathbf{K}(\mathbf{v} \text{ grad } v_z) = -(e/m)\mathbf{E}. \qquad \text{......(6)}$$

Furthermore, the following relations have to be satisfied:

$$\text{div}(\rho\mathbf{v}) = 0, \qquad \text{......(7)}$$

$$\text{div } \mathbf{E} = \rho/\epsilon_0, \qquad \text{......(8)}$$

$$\mathbf{E} = -\text{grad } V, \qquad \text{......(9)}$$

where ρ, V are respectively the charge density and the potential and ϵ_0 is the permittivity of free space (in rationalized M.K.S. units). These equations express respectively the continuity of current, Poisson's law and the irrotational nature of the electric field.

Equations (6), (7), (8) and (9) are equivalent to eight simultaneous equations in eight unknowns, namely, three components of velocity, three components of electric field, charge density and potential. Solutions can be obtained in the following way: choose as the values of the components v_x, v_y and v_z of the velocity **v** definite functions of x, y and z but containing arbitrary constant parameters. This will define a certain type of motion which will be a possible one if the arbitrary parameters can be adjusted so as to satisfy the conditions given below. Having thus chosen v_x, v_y and v_z, form the "acceleration vector" **A** given by

$$\mathbf{A} = \mathbf{I}(\mathbf{v} \text{ grad } v_x) + \mathbf{J}(\mathbf{v} \text{ grad } v_y) + \mathbf{K}(\mathbf{v} \text{ grad } v_z). \qquad \text{......(10)}$$

Then equation (9) will be satisfied provided

$$\text{curl } \mathbf{A} = 0, \qquad \text{......(11)}$$

and equation (7) will be satisfied provided

$$\text{div}(\mathbf{v} \text{ div } \mathbf{A}) = 0, \qquad \text{.......(12)}$$

while equations (6) and (8) then serve merely for calculating the field and charge density.

A has to satisfy these two conditions (11) and (12). If the arbitrary parameters can be chosen suitably, the resulting values of **v** and **A** define a possible space-charge motion.

In terms of **v** and **A** other important quantities of the motion are given by

$$\rho = (-m\epsilon_0/e) \text{ div } \mathbf{A}, \qquad \text{......(13)}$$

$$\mathbf{E} = (-m/e)\mathbf{A}, \qquad \text{......(14)}$$

$$V = (m/e)\int \mathbf{A}\,d\mathbf{s}, \qquad \text{......(15)}$$

$$\mathbf{i} = (-m\epsilon_0/e)\mathbf{v} \text{ div } \mathbf{A}, \qquad \text{......(16)}$$

where **i** is the current density vector.

Trajectories, total current and transit times are then also easily calculable, as will be seen in examples given below.

§ 3. SOLUTIONS OBTAINED

Example 1.

$$\mathbf{v} = a(\mathbf{I}x - \mathbf{J}y), \qquad \ldots\ldots(17)$$

$$\mathbf{A} = a^2(\mathbf{I}x + \mathbf{J}y). \qquad \ldots\ldots(18)$$

It is easily verified that equations (11) and (12) are satisfied. Quantities of the motion are obtained from equations (13), (14), (15) and (16):

$$\rho = -2a^2(m\epsilon_0/e), \qquad \ldots\ldots(19)$$

$$\mathbf{E} = -a^2(m/e)(\mathbf{I}x + \mathbf{J}y), \qquad \ldots\ldots(20)$$

$$V = a^2(m/2e)(x^2 + y^2) + \text{constant}, \qquad \ldots\ldots(21)$$

$$\mathbf{i} = -2a^3(m\epsilon_0/e)(\mathbf{I}x - \mathbf{J}y). \qquad \ldots\ldots(22)$$

It follows from equation (17) that the trajectories are given by the equation $dy/dx = -y/x$, i.e. $xy = \text{constant}$. Thus the trajectories are rectangular hyperbolae in planes perpendicular to the z axis and asymptotic to the diametral planes $x = 0$ and $y = 0$.

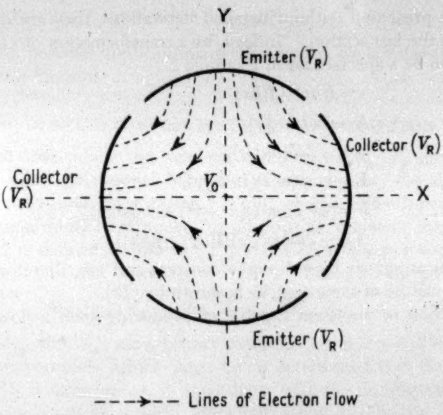


Figure 1. The motion $\mathbf{v} = a(\mathbf{I}x - \mathbf{J}y)$.

A cross section of a system in which such a flow will take place is shown in Figure 1. It consists of an infinitely long conducting cylinder of radius R and potential V_R, the axis of which coincides with the z axis. The cylinder is divided into four quadrants by cuts parallel to the axis; two opposite quadrants constitute emitters of electrons and the remaining two collectors. The potential on the axis V_0, follows from equation (21):

$$V_0 = V_R - a^2(m/2e)R^2. \qquad \ldots\ldots(23)$$

This former equation also shows that the equipotentials are concentric cylinders.

The total current I_{tot} per unit length of the system is obtained by making use of equation (22) to integrate the current flow over the area of unit length of the two emitters:

$$I_{\text{tot}} = 4a^3(m\epsilon_0/e)R^2. \qquad \ldots\ldots(24)$$

From equations (23) and (24) follow the relations between current and potential difference:

$$I_{\text{tot}} = 8\epsilon_0(2e/m)^{\frac{1}{2}}(1/R)(V_R - V_0)^{\frac{3}{2}}. \qquad \ldots\ldots(25)$$

The transit time τ along a trajectory $xy = \pm\frac{1}{2}r_0^2$ (where r_0 represents the distance of closest approach of the trajectory to the axis) is given by

$$\tau = (1/a)\cosh^{-1}(R^2/r_0^2). \qquad \ldots\ldots(26)$$

The interesting result is thus obtained of a flow with transit times varying from zero (for $r_0 = R$) to infinity (for $r_0 = 0$). Finally, it may be noticed, as is seen in equation (19), that the charge density is everywhere constant.

This example has been presented primarily for its theoretical interest. To achieve such a system in practice would be difficult, chiefly because of the awkward directions of entry of the electrons at the emitters. But an approximation to it involving the use of grids as "emitters" with electrons entering from outside the system might find applications in space-charge focusing (constant charge density), wide-band ultra-high-frequency oscillators (large range of transit times) and 90° beam deflections without aid of a magnetic field.

Example 2.

Results are presented without detailed derivation; they are deduced in the same way as in the last section. Indeed, by a transformation of axes, Example 1 can be shown to be a special case of Example 2.

$$\mathbf{v} = \mathbf{I}(ay) + \mathbf{J}(bx), \qquad \ldots\ldots(27)$$

$$\mathbf{A} = ab(\mathbf{I}x + \mathbf{J}y), \qquad \ldots\ldots(28)$$

$$\rho = -2ab(m\epsilon_0/e), \qquad \ldots\ldots(29)$$

$$\mathbf{E} = -ab(m/e)(\mathbf{I}x + \mathbf{J}y), \qquad \ldots\ldots(30)$$

$$V = (ab/2)(m/e)(x^2 + y^2) + \text{constant}, \qquad \ldots\ldots(31)$$

$$\mathbf{i} = -2ab(m\epsilon_0/e)\{\mathbf{I}(ay) + \mathbf{J}(bx)\}. \qquad \ldots\ldots(32)$$

The trajectories are given by $bx^2 - ay^2 = \text{constant}$ and are, therefore, hyperbolae since a and b must be of same sign, as is shown by (29).

A cross-section of a system capable of producing such a flow is shown in Figure 2.

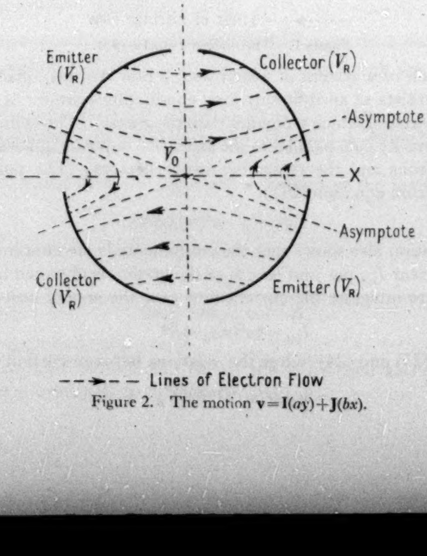


Figure 2. The motion $\mathbf{v} = \mathbf{I}(ay) + \mathbf{J}(bx)$.

It is clear that the trajectories are now asymptotic to the diametral planes

$$x\sqrt{b}-y\sqrt{a}=0; \qquad x\sqrt{b}+y\sqrt{a}=0. \qquad \ldots\ldots(33)$$

The potential on the axis V_0 is given by $V_0=V_R-ab(m/2e)R^2$, where V_R and R represent the potential and radius of the electrodes. The equipotentials are again concentric cylinders.

The total current I_{tot} is given by $I_{\text{tot}}=2ab(a+b)(m\epsilon_0/e)R^2$. The total current is proportional to the three-halves power of potential difference V_R-V_0, but the relation cannot be put in as perspicuous a form as that of equation (25).

For a trajectory $bx^2-ay^2=c$ (c positive) passing through the $y=0$ diametral plane the transit time τ is given by

$$\tau=2\{\sqrt{(ab)}\}^{-1}\cosh^{-1}\{\sqrt{[b(aR^2+c)/c(a+b)]}\}, \qquad \ldots\ldots(34)$$

and for a trajectory $bx^2-ay^2=-c$ (c positive), passing through the $x=0$ diametral plane (Figure 2), by

$$\tau=2\{\sqrt{(ab)}\}^{-1}\cosh^{-1}\{\sqrt{[a(bR^2+c)/c(a+b)]}\}. \qquad \ldots\ldots(35)$$

Comments on possible applications of this system are similar to those given in connection with Example 1.

§4. DISCUSSION OF METHOD OF SOLUTION

Consider the case when the velocity components v_x, v_y and v_z may be expressed as integral polynomial functions of x, y and z; the dummy suffix and Kronecker delta notation are used. Thus, when a suffix appears twice in a nomial, that nomial must be summed for all possible values of the suffix, while the symbol, δ^{lmn}_{pqr} is defined as follows: $\delta^{lmn}_{pqr}=1$ if $l=p$, $m=q$, $n=r$, and $\delta^{lmn}_{pqr}=0$ in all other cases. Further, write $S_{lmn}=x^l y^m z^n$. Then, for the polynomial case

$$\mathbf{v}=\mathbf{I}a_{lmn}S_{lmn}+\mathbf{J}b_{lmn}S_{lmn}+\mathbf{K}c_{lmn}S_{lmn}, \qquad \ldots\ldots(36)$$

where the a_{lmn}, b_{lmn} and c_{lmn} are arbitrary coefficients and l, m and n are arbitrary indices of the polynomials, which must be so determined that the motion is a possible one. [It is to be noted, in accordance with the dummy suffix notation, that an expression such as $a_{lmn}S_{lmn}$ does not mean merely the single nomial $a_{lmn}x^l y^m z^n$, but the sum of terms like it, for all possible values of l, m and n, viz. $\Sigma_l\Sigma_m\Sigma_n a_{lmn}x^l y^m z^n$. The values of l, m and n need not necessarily be integral.]

Hence, from equation (10),

$$\mathbf{A}=T_{lmn}(\mathbf{I}a_{lmn}+\mathbf{J}b_{lmn}+\mathbf{K}c_{lmn}),$$

where T_{lmn} is given by

$$T_{lmn}=S_{l'm'n'}S_{lmn}[(l/x)a_{l'm'n'}+(m/y)b_{l'm'n'}+(n/z)c_{l'm'n'}].$$

The two fundamental conditions expressed by equations (11) and (12) can now be written as follows:

Condition 1.

$$\partial/\partial x(b_{lmn}T_{lmn})=\partial/\partial y(a_{lmn}T_{lmn}),$$

$$\partial/\partial y(c_{lmn}T_{lmn})=\partial/\partial z(b_{lmn}T_{lmn}),$$

$$\partial/\partial z(a_{lmn}T_{lmn})=\partial/\partial x(c_{lmn}T_{lmn}).$$

Condition 2.

$$\partial/\partial x(a_{l'm'n}S_{l'm'n'}W)+\partial/\partial y(b_{l'm'n'}S_{l'm'n'}W)+\partial/\partial z(c_{l'm'n'}S_{l'm'n'}W)=0,$$

where W is given by

$$W=\partial/\partial x(a_{lmn}T_{lmn})+\partial/\partial y(b_{lmn}T_{lmn})+\partial/\partial z(c_{lmn}T_{lmn}).$$

Particular results are deduced from the above formulation of the problem in the following manner.

Case 1.

Assume $b_{lmn}=0;\quad c_{lmn}=0;\quad a_{lmn}=a\delta^{lmn}_{pqr},$

i.e. assume that the flow is entirely in the direction of the x axis and the velocity is a monomial function of the coordinates, since only when l, m and n have the particular values p, q and r respectively is the coefficient of $x^l y^m z^n$ not zero; the application of the two fundamental conditions will show what magnitudes p, q and r must have for a possible flow.

Condition 1 shows that $q=0$, $r=0$, while Condition 2 leads to the equation $p(2p-1)(3p-2)=0$ for p. By referring to equation (13) it is seen that the solutions $p=0$, $p=\frac{1}{2}$ both give possible motions involving zero charge density and, therefore, of no present interest.

The solution $p=\frac{2}{3}$ gives the Child–Langmuir flow between infinite parallel planes and, for purposes of reference, we give here the characteristics of the flow, deducible from equations (13) to (16):

$$\mathbf{v}=\mathbf{I}ax^{\frac{2}{3}};\qquad \mathbf{A}=\mathbf{I}(2a^2/3)x^{\frac{1}{3}};$$

$$\rho=(-2a^2/9)(m\epsilon_0/e)x^{-\frac{2}{3}};\qquad \mathbf{E}=\mathbf{I}(-2a^2/3)(m/e)x^{\frac{1}{3}};$$

$$V=(a^2/2)(m/e)x^{\frac{4}{3}};\qquad \mathbf{i}=\mathbf{I}(2a^3/9)(m\epsilon_0/e).$$

Case 2.

Assume $c_{lmn}=0;\quad a_{lmn}=a\delta^{lmn}_{pqr};\quad b_{lmn}=b\delta^{lmn}_{p'q'r'},$

i.e. that there is no component of flow in the direction of the z axis, while the other two components are monomial functions of the coordinates.

By applying Conditions 1 and 2 it is found that, essentially, only two solutions exist involving non-zero charge density, namely,

$$p=1,\quad q=0,\quad r=0,$$
$$p'=0,\quad q'=1,\quad r'=0,$$
$$a=-b,$$

which leads to the motion studied in § 3, Example 1, and

$$p=0,\quad q=1,\quad r=0,$$
$$p'=1,\quad q'=0,\quad r=0,$$
$$a,\ b \text{ of same sign}$$

which leads to the motion studied in § 3, Example 2.

It is seen that these two examples represent two of the simplest cases of a limited (integral polynomial) class of solutions. By considering binomials and other polynomials, and the other classes of function (e.g. fractional polynomials, trigonometric functions, hyperbolic functions, etc.), a vast field of knowledge can be opened up. It is to be noted in passing that solutions discarded because they involve zero charge-density are still valid solutions and may be of considerable interest in low-current or high-voltage phenomena.

Finally, it must be pointed out that there is considerable scope for pure mathematicians in building up existence and uniqueness theorems, establishing criteria for solutions with given boundary conditions, studying effects of co-ordinate transformations and, in fact, building up a body of theory for electron flow comparable to that of potential theory.

REFERENCES

Bellustin, S. V., 1939, *J. Phys. U.S.S.R.*, I, **3**, 251.
Langmuir, I., 1913, *Phys. Rev.*, **2**, 450.
Langmuir, I., and Blodgett, K. B., 1924, *Phys. Rev.*, **24**, 49.

The Accuracy of Measurements by Rayleigh Disc

By W. WEST
Post Office Engineering Research Station, Dollis Hill, London

Communicated by Physical Society Acoustics Group; *MS. received 25th November* 1948

ABSTRACT. A review of proposals made for altering the formula due to König, on which are based measurements of air-particle velocity by Rayleigh disc, leads to the conclusion that there is insufficient evidence for altering the numerical constant in the formula. Comparisons between Rayleigh disc and other methods of calibration of individual microphones are discussed, and the possibility of difference due to reaction on the sound pressure by the vibration of the diaphragm is considered.

§1. DIRECT EVIDENCE (MEASUREMENTS OF PARTICLE VELOCITY)

The familiar formula established by König (1891) from hydrodynamical theory is now generally accepted as the basis for measurements of sound particle velocities, and hence of sound pressures, by the Rayleigh disc. The formula is calculated for an ellipsoid in an infinitely extended, uniformly moving, non-viscous fluid, and is therefore open to corrections arising from practical departures from these ideal conditions. The importance of the Rayleigh disc to acoustical measurements has led several investigators to determine, by experiment, the magnitude of the small corrections to be applied for such of the departures from the ideal as are inevitably encountered in the use of the Rayleigh disc for calibrating microphones. In an account of the most recent of these investigations Scott (1945) gives a valuable summary of information concerning these corrections—to which his own investigation contributes very substantially; but for a correction of the numerical constant in the formula proposed by Merrington and Oatley (1939), and to some extent supported by Scott, no explanation has been given. The main purpose of the present article is to examine the justification for applying this correction.

König's formula, as applied to very thin circular discs, states that

$$L = \tfrac{4}{3}\rho a^3 V^2 \sin 2\theta, \qquad \ldots\ldots(1)$$

where L is the torque on a disc of radius a, set in a stream of fluid of density ρ and velocity V, so that the normal to the disc makes angle θ with the direction of flow.

Wood (1935) has drawn attention to the fact that in König's formula the velocity V is the relative velocity between the fluid and the disc, whereas with a freely suspended disc (as ordinarily used) an alternating movement of the fluid imparts a movement to the disc. With sufficient accuracy for the present purpose, the effect can be allowed for by introducing the factor $(1-8a^3\rho/3M)$ on the right-hand side of equation (1), provided that the quantity $8a^3\rho/3M$, where M is the mass of the disc, is small as compared with unity. This correction has a sound theoretical basis (Wood 1935, King 1935), and has received substantial experimental verification by Wood from experiments made in water, for which, due to the greater density, the effect is very much more pronounced than in air.

Merrington and Oatley (1939) proposed a modification of the formula, which not only includes a correction factor for this effect but also alters the numerical constant in the formula from $\frac{4}{3}$ to 1·47 (an increase of about 10%). No theoretical justification is made for this change, and it is based solely on the results of their experimental work. No correction approaching this magnitude was found by earlier investigators (Zernov 1908, Barnes and West 1927), who had used essentially the same technique, namely, applying the air particle velocity by vibrating, at low frequency, a box surrounding the disc and by measuring the amplitude of vibration of the box by microscope.

Zernov has given a reasonably full and complete account of his experimental method and results, as far as his tests with ellipsoids are concerned. He concludes from his experiments that König's formula is fully confirmed for ellipsoids in air to within an experimental margin of error of 0·5%. In arriving at his result, however, Zernov introduced a purely empirical correction, amounting to 2·5%, for the effect of the proximity of the walls of the (cylindrical) vibrating box to the ellipsoid. He quotes a formula for this correction, but gives no information of the data on which he based the formula. The value of correction (2·5%) which he used is in the same sense, but is about twice as large as that found by Scott (1945, Figure 3). Zernov extended his investigation to discs and found discrepancies (due to finite thickness of disc) for which he proposed an empirical correction, without quoting the data on which it was founded. This correction is also in the same sense, but greater than that found by Scott (1945, Figure 4); it is, however, in the opposite sense to that calculated by König for a non-viscous fluid. The ellipsoids and discs used by Zernov were massive (on modern standards, and Wood's correction mentioned above would be quite negligible.

The absence of data given by Zernov for his empirical corrections should not be allowed to detract from the results of the work on which most of his article is concentrated, namely, the experimental verification of König's formula for ellipsoids. Even if the empirical correction which he made for the proximity of the walls of the vibrating box is completely ignored, his experimental results lie within 3% of the value given by theory. From his experiments with discs he deduced that, for very thin discs, any correction for alteration of shape from ellipsoid to flat disc is very small, that the correction is increased by increasing the thickness of the disc, and that it has the opposite sense to that calculated by König for non-viscous fluids; these deductions have received general confirmation from subsequent investigation (Scott 1945).

From their experiments Barnes and West (1927) concluded that, "provided the conditions of use conform within reasonable limits with those on which König's formula is based, then the measurement of acoustic velocities, at any rate down to the order of 0·1 cm/sec., can be effected by means of a Rayleigh disc to an accuracy within 2%" (which corresponds to accuracy within 4% for the torque on the disc). Merrington and Oatley discarded the evidence from the experiments by Barnes and West on the grounds that two omissions were made in their calculations. One was the correction factor $(1-C_1)$ named above, where $C_1 = 8a^3\rho/3M$. The other was the omission of the increment to inertia due to the air in the calculation of the moment of torsion of the suspension fibre; it can also (for convenience of this discussion) be resolved into a correction factor $(1-C_2)$ on the right-hand side of equation (1) where C_2 is the ratio $I_0 : I$ of the moment of inertia due to the air to that due to the disc alone. Now $I_0 = 16\rho a^5/45$ (Lamb 1906) and $I = Ma^2/4$, whence $C_2 = 64\rho a^3/45M$.

Now information from which the values of C_1 and C_2 can be calculated is given by Barnes and West for two of the discs used by them for their fundamental calibration, as follows:

	M (gm.)	a (cm.)	C_1	C_2
First disc	0·0764	0·6115	0·0096	0·005
Second disc	0·0622	0·473	0·0055	0·003

Thus these two corrections, even when combined, amount to appreciably less than 2%.

Even if all evidence other than that presented by Merrington and Oatley is discarded it is doubtful whether their experiments were sufficiently complete to justify an empirical correction of 10% to the theoretical formula. There is one aspect of their measurement of the air-particle velocity which does not seem to have received the attention it requires, and that is the wave-form of the displacement of the box (and hence of the air) surrounding the disc. The measurement of the quantity V^2 in König's formula was derived from measurements of frequency and of maximum amplitude of the displacement of the box. Now the presence of a small harmonic component in the wave-form can not only add to the mean-square value V^2, but it can also alter the maximum amplitude of the wave.

Take, for example, the case of a sine-wave with maximum amplitude of displacement A and angular frequency ω; the wave-form of displacement is $A \sin \omega t$ and the mean-square velocity is $A^2\omega^2$. If a small component of third harmonic is added, with such phase as to reduce the maximum amplitude, the wave-form of displacement having maximum amplitude A is

$$A(1+x) \sin \omega t + Ax \sin 3\omega t$$

where x is the third harmonic fraction of the fundamental. The mean-square velocity $A^2(1+x)^2\omega^2 + 9A^2x^2\omega^2$ is greater than that of the pure sine-wave in the ratio $(1+x)^2 + 9x^2 : 1$. Thus, for example, 4% of third harmonic $(x = 0{\cdot}04)$ can introduce an error in measurement of V^2 of nearly 10%. Higher order harmonics can, of course, introduce relatively greater errors, for example the error with 2% of 11th harmonic could be nearly 9%.

Now while there may not be sufficient information for asserting that a small distortion of wave-form did in fact contribute to the discrepancies between König's formula and Merrington and Oatley's experiments, there is too much room for doubt to permit acceptance of their proposal to alter the numerical constant in the formula. In Table II of their paper specific amplitudes of vibration of the box are stated, and it is quite consistent that, for tests on any one disc, the discrepancy between experiment and calculation does increase quite markedly with increase in amplitude. The mechanism which they used for vibrating the box (enclosing the disc) included a valve amplifier and an electromagnetic drive—a method which is more open to distortion of wave-form than the simple and controlled form of mechanical drive used by Barnes and West.

At a later date Scott carried out some experiments by a different technique which enabled measurements to be made at frequencies up to 4,000 cycles per second. He, too, has found discrepancies with König's formula, though of less magnitude than those observed by Merrington and Oatley. His comment that the discrepancies which he has observed are too large to be accountable in terms of experimental uncertainty is not thought to be sound. It leads to the curious conclusion that the numerical constant ($\frac{4}{3}$) in König's formula is too small by about 7% at low frequencies and by about 4% at high frequencies (2,500 to 4,000 c/s.), but is correct at frequencies between about 1,000 and 2,000 c/s., irrespective of the diameter of the disc.

Criticism of these two articles is levelled only at the particular conclusion, reached in both of them, that the experimental uncertainty is sufficiently small to detect and correct for a departure from the theoretical formula, even though no theoretical explanation of such a departure can be put forward and established. There is much other matter which is of value to the art of acoustical measurements in both articles.

It is reasonable to conclude that, in the interests of accuracy of measurement, it is advisable to adopt a technique in which any corrections to be applied to König's formula are reduced to very small values. Scott's paper, in particular, contains much useful practical information for guidance on the probable magnitude of corrections arising from necessary practical departures from conditions assumed in the theory.

The correction introduced by Wood, i.e. the factor $(1-C_1)$ mentioned above, is a valuable modification of König's formula for application to the calibration of a Rayleigh disc. For measurements of sound in air it is convenient to use a disc having small moment of inertia, so that the decay factor is not too small and the time required for taking readings is not unduly long. A strip of relatively massive material, fixed along the vertical diameter of the Rayleigh disc, can increase very substantially the mass (and thereby reduce the amount of the correction) without appreciably altering either the moment of inertia of the disc or the validity of König's formula. This device is mentioned in § 2 of the paper by Barnes and West.

§ 2. INDIRECT EVIDENCE (MEASUREMENTS OF SOUND PRESSURE)

The principal practical application of the Rayleigh disc is to the calibration of stable, high-quality microphones for use for the measurement of sound pressures. Reference will therefore be made in this section to some comparison tests which have been made between calibrations of individual microphones by Rayleigh disc and by other methods. Percentage differences which are quoted here apply to sensitivities of microphones (that is, the ratio of E.M.F. to sound pressure) and not to torque on the Rayleigh disc as in the previous section.

As a means of practical verification of König's formula to any high degree of accuracy, such tests are less direct and more open to experimental errors than those discussed in § 1. Uncertainties are introduced, not only by the accuracy limitations of the compared method of measuring the sound pressure applied to the microphone, but also in the accurate measurement of the small E.M.F. produced by the microphone, in the dependence on stability of performance by the microphone over the interval of time between the two calibrations compared, and by possible effects on the sensitivity of the microphone, due to differences between the acoustical conditions in which it is used for the calibrations which are being compared. (A possible example of an effect of this kind is mentioned later.) None the less, the evidence which has been obtained from comparisons of calibrations by different methods has relevance and interest.

Dadson (1946) reports comparisons made by the National Physical Laboratory of calibrations of a condenser microphone by Rayleigh disc and by pistonphone over the frequency range 60 to 400 c/s.; the agreement was within 3%, the estimated accuracy of the pistonphone method being to within 3 or 4%.

Field calibrations of a microphone by pistonphone and by Rayleigh disc have been compared in the frequency range 50 to 1,000 c/s. by Kennedy and Boner (1942). They found substantial agreement at 50 c/s. and between 170 and 1,000 c/s., maximum differences being less than 10% and the average of differences practically zero. Between 50 and 170 c/s. there were greater differences attributed by the authors to "wobble" of the piston.

For his investigations of effects of obstruction and cavity resonance, Ballantine (1930, 1932) measured pressure calibrations by electrostatic actuator, checked by thermophone, and field calibrations by Rayleigh disc. At low frequencies, where these effects are shown to be small, all the points on his graphs of the ratio of these two measurements lie within 2 or 3% of unity.

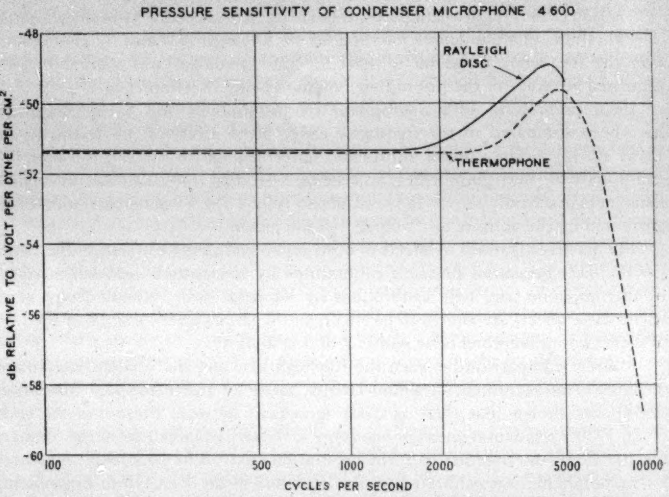
A direct comparison between the Rayleigh disc and the acoustic radiometer methods of measurement of sound energy, made by Hartmann and Mortensen (1948), has shown that there is close agreement between these two methods.

In 1930 a condenser microphone (type D 85880), obtained from the Western Electric Company together with the calibration chart of its sensitivity, measured by thermophone, was calibrated by Rayleigh disc at the Post Office Engineering Research Station. The results of this comparison have not hitherto been published. The Rayleigh disc was used to measure both the field sensitivity and the pressure sensitivity of the microphone.* The Figure shows the results of the two pressure calibrations, by thermophone as supplied by the makers and by Rayleigh disc as measured after delivery. It happens that at the lower frequencies, below about 1,500 c/s., the two curves coincide; that for the thermophone test is level at 0·0027 v/dyne/cm^2, and seven points in this range measured by Rayleigh disc all lie between 0·00265 and 0·00275. At higher frequencies the Rayleigh disc curve tends to depart by up to about 15% from the thermophone curve, and it appears to reach a maximum sensitivity at a somewhat lower frequency.

It may well be that the differences lie within the combined experimental errors of the two tests, plus a possible change in the performance of the microphone between the tests. There is, however, also the possibility of a slight difference in shape of the frequency characteristics caused by the stiffness reactance imposed on the diaphragm by the volume of about 9 cm^3 of hydrogen at atmospheric pressure in the thermophone chamber. This stiffness adds to that of the diaphragm and to that due to the enclosed air-space behind the diaphragm, and it tends to raise the frequency of maximum response associated with the highly-damped mechanical resonance of the diaphragm. The argument is speculative in the absence of quantitative knowledge of the mechanical impedance of the diaphragm and of the air space behind. This type of condenser microphone has been in use in this country and elsewhere as a component of a standard reference system for telephone transmission tests; repeated calibration tests carried out over a period of many years have shown that individual microphones have a stability which is of a very high order.

---

* The methods and apparatus used for the calibrations are described by West (1932). The method used for measurement of field sensitivity has since been changed at this Research Station; the pressure sensitivity of a probe microphone (West 1934), which has a reasonably uniform response–frequency characteristic, is first measured by Rayleigh disc in the resonant tube; this microphone is then used to measure the sound field from a loudspeaker in a relatively large non-reverberant room. (As an obstruction to the sound field the probe is no larger, and can well be smaller, than the Rayleigh disc and it is much more convenient to use in the open.) The microphone under test is then measured at the point where the sound field has been calibrated by the probe microphone.



§ 3. POSSIBLE EFFECT ON CALIBRATION OF THE ACOUSTICAL IMPEDANCE TO WHICH THE DIAPHRAGM IS EXPOSED

As a general comment on the calibration of microphones it is of interest to draw attention to the possibility of difference between results for different kinds of pressure calibration (particularly at frequencies where the response is influenced by mechanical resonance of the diaphragm) due to differences in the acoustical impedances to which the diaphragm is exposed in the different measurements, if the reaction of the vibration of the diaphragm on the acoustical impedance is ignored.

With a true pressure calibration—where measurement is made of the pressure existing at the surface of the diaphragm—the acoustical impedance in front of the diaphragm exerts no influence. The author felt the need for experimental verification of this fact, and some calibrations were made at frequencies near resonance of a telephone receiver, used alternatively as a microphone and as a source of sound, with different acoustical impedances imposed by forming an enclosure in front of the telephone receiver and by altering the volume of this enclosure. A calibrated probe microphone was used for measuring the sound pressures at a point just in front of the diaphragm. A source of sound formed part of the wall of the enclosure for generating sound pressures therein when the telephone receiver was being calibrated as a microphone. Electrical measurements were made in terms of voltages at the receiver terminals. The mechanical resonance of the diaphragm showed, as expected, a very pronounced peak in all the tests, but whereas the frequency of maximum sensitivity changed from about 1,300 to about 1,000 c/s. for variations of enclosure volume from 1·5 to 20 cm^3 when the telephone receiver was calibrated as a source of sound, it remained constant at about 1,000 c/s. when the telephone receiver was calibrated as a microphone. (Precautions were taken to eliminate changes due to temperature variations.)

For the microphone calibration, the pressure in the enclosure may be regarded as made up of two components, one causing the diaphragm to vibrate and the other resulting from its vibration. The relative magnitudes and phases of these components alter as the frequency passes through that of resonance of the diaphragm. It would appear that the analogue, using the impedance concept applied to electrical circuits, is that of a generator, having an internal reactive impedance, applied to a series-tuned circuit (the diaphragm); the response of the tuned circuit, in terms of the ratio of current flowing in it to the potential difference across its terminals (i.e. ratio of velocity of diaphragm to pressure in enclosure), is independent of the impedance of the generator. The two components of sound pressure in the enclosure are represented, respectively, by the E.M.F. and the potential difference across the internal impedance of the generator. The corresponding analogue in the case of the calibration as a source of sound would be to regard the E.M.F. as generated in the tuned circuit (which has the reactive impedance of the enclosure in series with it) and the response of the mechanical system as the ratio of the potential difference across the impedance of the enclosure to the E.M.F. (i.e. of pressure in enclosure to force on diaphragm).

Some of the methods used for measuring pressure sensitivities of microphones do not yield true pressure calibrations (in the sense used here). Thus, for example, in the thermophone and the electrostatic actuator methods it is usual to ignore the contribution to the pressure on the diaphragm attributable to the movement of the diaphragm, and, therefore, to include the effect of the acoustical impedance to which the diaphragm is exposed in the calibration. It is true that at frequencies remote from that of resonance of the diaphragm (or in cases where the resonance is so highly damped as to exert no appreciable influence on the response) the contribution to the pressure due to the movement of the diaphragm is very small; in other words the load imposed by the acoustical impedance is an insignificant fraction of the mechanical impedance of the diaphragm. In fact the effect has been considered and quantitative estimates of its magnitude made by Ballantine (1932) for the thermophone method of calibration, but consideration was limited to frequencies at which the diaphragm is stiffness-controlled: it did not include the increase of magnitude and change of phase which occurs at resonance of the diaphragm.

The Rayleigh disc method of calibration in a resonant tube does yield a true pressure calibration, since any contribution to the stationary wave in the tube caused by vibration of the diaphragm is included in the measurement made by the Rayleigh disc. Thus, if mechanical resonance of the diaphragm does exert appreciable influence on the response of a microphone in the range of frequencies tested, it is to be expected that the maximum response due to the resonance will occur at a somewhat higher frequency for the thermophone than for the Rayleigh disc calibration. Moreover, the sensitivity at the point of maximum response depends largely on the amount of mechanical resistance opposing the movement of the diaphragm (in the case of a condenser microphone the resistance resides mostly in the small air space between the electrodes). Since the resistance tends to increase as the frequency is raised, a somewhat smaller maximum response is to be expected when the maximum occurs at a higher frequency. The result is a difference (though possibly only slight) in the shape of the frequency-response characteristic as measured by these two methods of calibration.

In the case of a measurement of field sensitivity, the sound pressure measured is that existing in the free air at the position of, but in the absence of, the microphone. There is thus no question of movement of the diaphragm contributing to the value of the pressure measured, and the acoustical impedance load on the diaphragm is that due to the free air—mostly a small inertia reactance.

ACKNOWLEDGMENT

The author is indebted to Mr. J. O. Ackroyd for much helpful comment and criticism in the preparation of this paper.

REFERENCES

Ballantine, S., 1930, *Proc. Inst. Radio Engrs.*, **18,** 1206; 1932, *J. Acoust. Soc. Amer.*, **3,** 319.

Barnes, E. J., and West, W., 1927, *J. Instn. Elect. Engrs.*, **65,** 871.

Dadson, R. S., 1946, *Proc. Roy. Soc.* A, **186,** 207.

Hartmann, J., and Mortensen, T., 1948, *Phil. Mag.*, **39,** 377.

Kennedy, W. J., and Boner, C. P., 1942, *J. Acoust. Soc. Amer.*, **14,** 19.

King, L. V., 1935, *Proc. Roy. Soc.* A, **153,** 17.

König, W., 1891, *Ann. Phys., Lpz.*, **43,** 43.

Lamb, H., 1906, *Hydrodynamics* (Cambridge: University Press), p. 140.

Merrington, A. C., and Oatley, C. W., 1939, *Proc. Roy. Soc.* A, **171,** 505.

Scott, R. A., 1945, *Proc. Roy. Soc.* A, **183,** 296.

West, W., 1930, *J. Instn. Elect. Engrs.*, **68,** 441; 1932, *Acoustical Engineering* (London: Pitman), §§ 59, 60; 1934, *P.O. Elect. Engrs. J.*, **26;** 260.

Wood, A. B., 1935, *Proc. Phys. Soc.*, **47,** 779.

Zernov, W., 1908, *Ann. Phys., Lpz.*, **26,** 79.

A Dynamic Gravimeter of Novel Design*

By R. L. G. GILBERT
Department of Geodesy and Geophysics, Cambridge

Communicated by B. C. Browne; MS. received 14th January 1949

ABSTRACT. A new type of gravimeter is described, which depends for its operation on the change of frequency of the natural vibration of a vertical wire stretched by a weight. The advantages of such an instrument are presented, and a description given of the design which has so far been evolved. The results of laboratory tests and of trials carried out at sea in H.M. Submarine *Talent* are given.

---

§1. INTRODUCTION

At present the pendulum is the "standard" instrument for relative measurements of gravity: all other gravimeters have to be checked against this standard more or less frequently. This is usually effected by visiting known gravity stations at intervals, and working thereafter in a closed loop each day, so as to allow for the daily drift of the instrument.

One of the aims in making the vibration type of gravimeter was to provide a substandard to the pendulums. The cause of drift in nearly all other existing gravimeters lies in the spring: the extension of the spring is directly proportional to the total attraction of gravity. Since the elastic properties of most materials vary slowly with time, the zero reading of the gravimeter will alter, and this limits the time over which the instrument may be used to obtain results accurate to any given amount. It is possible, however, to design a gravimeter in which the elastic properties of the materials do not affect the readings directly. Such an instrument has been described by Bertrand (1938). It consists of a mass suspended from a fixed support by a thin wire. If the wire is made to vibrate transversely, its natural period will depend on the tension, and hence on the value of gravity.

The frequency of the real wire differs from that of an ideal, perfectly flexible, inextensible wire for three reasons: firstly, the wire has a finite rigidity; secondly, the length of the wire varies with the tension; and thirdly, the supports at each end of the wire are not absolutely rigid. Suitable design renders these terms small, and drift due to variations in them can be largely eliminated. The frequency of the ideal string depends inversely on its length and the square root of its mass per unit length and directly on the square root of the tension. The tension in the string depends on the mass suspended by it, and this and the mass of the string can easily be kept constant to 1 in 10^7 provided precautions are taken to avoid corrosion. The length of the string, apart from the effect of creep, depends only on the temperature, and the apparatus is accordingly placed in a thermostat. Experiment shows that the effect of creep can be reduced to less than 1 part in 10^6 per day. The frequency is also dependent on the air pressure, as in the case of pendulums, and the apparatus is therefore evacuated.

The principal advantage which the instrument possesses over an ordinary gravimeter is its extended scale. The time needed for an observation to attain a desired degree of accuracy depends only on the accuracy with which the phase of the wire can be measured with respect to a fixed frequency standard. At present this is about 1/10 cycle and, since the string vibrates at 1,000 c/s., a reading correct to 1 in 10^6 can be obtained in 100 seconds. This accuracy is unaffected by the actual frequency difference between the wire and the standard; thus variations in g of 2%, such as are encountered over the earth, can be measured with the same accuracy. The instrument is therefore suitable for geodetic use.

---

* Patents pending.

The accuracy of phase-comparison could easily be improved to 1/50 cycle and thus an accuracy of reading of 1 in 10^8 obtained in just over half an hour. The instrument could then be used to measure earth tides, provided that its drift were uniform.

The instrument also possesses integrating properties. Since the beat-frequency is, to the first order, proportional to g, then the total number of beats measured in any given interval is proportional to the mean value of g, whatever may have been the form of its variation during the interval. This property makes the vibration gravimeter especially valuable for measurements in submarines. Indeed the only device which shares this property is a pendulum-apparatus, which so far has formed the only successful method of measuring gravity at sea.

The instrument is unlikely to compete with static gravimeters for ordinary commercial surveys, since its associated electronics and power supplies render it inconvenient. Nevertheless, it might be possible to make a model small enough to be lowered down a borehole. A well-known strain gauge depends on the principle outlined, but is not nearly so accurate as the instrument here described.

§ 2. DESIGN

In his original paper, Bertrand described the apparatus as follows:

"A vertical string is tensioned by the weight Mg of a mass M. The frequency of its fundamental $F=(1/2L)\sqrt{(Mg/m)}$ (L is the length of the string, m its mass per unit length) varies with g and enables variations therein to be measured. From the measured frequency one can deduce the value of g, using a constant applicable to the apparatus, which constant can be found by adding known weights to M."

A schematic diagram of the instrument is shown in Figure 1. The weight is suspended from a beryllium copper strip about 5 cm. long. This particular instrument was designed for use at sea, and it was therefore desirable that the weight hanging on the string should be damped, so as to prevent any pendulous swinging, with consequent inaccuracy in the determined value of gravity. The weight was made in the form of a copper disc, with a downward projecting rim, and a boss in the centre in which the half-cylinder at the end of the string could be clamped. A damping magnet was made, with four sets of poles symmetrically placed round the weight, which produced four horizontal radial fields, cutting the rim of the weight, and four vertical fields, cutting the disc portion of the weight. The latter was thus damped for lateral movements, twists about the axis of the disc, twists about any radius of the disc, and vertical movements. The constancy of the mass of the weight is assured, for copper does not readily corrode and has a low vapour pressure.

The string must have a cross-sectional area sufficient to support the weight; a tension of approximately one-fifth of the breaking stress is used. To the first order, the frequency of a flexible bar, of finite rigidity, clamped at each end, is given by Rayleigh (1929) as

$$n = \pi a/l[1 + 2(b/a)(\kappa/l)],$$

so long as $2(b/a)(\kappa/l)$ is small.

Here $a = \sqrt{(T/\rho)}$; $T\omega$ is the tension in the string, $\rho\omega$ the linear density, ω the area of cross section of the string, l the length, $b = \sqrt{[(q+T)/\rho]}$; q is the Young's modulus of the string, and κ is the radius of gyration of the section of the string about its median plane. $n = \pi a/l$ is, of course, the frequency of the ideal string.

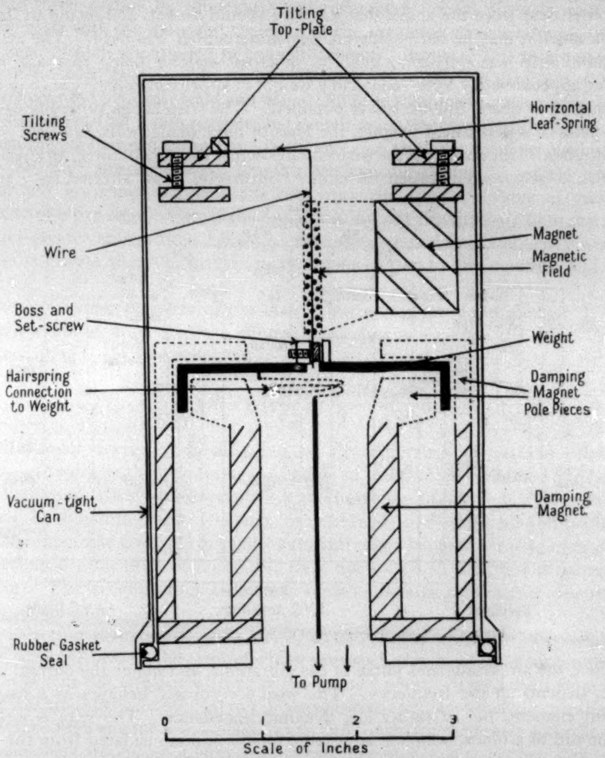

Tilting Top-Plate
Tilting Screws
Horizontal Leaf-Spring
Wire
Magnet
Magnetic Field
Boss and Set-screw
Weight
Hairspring Connection to Weight
Damping Magnet Pole Pieces
Vacuum-tight Can
Damping Magnet
Rubber Gasket Seal
To Pump
0
1
2
3
Scale of Inches

Figure 1. Schematic diagram of gravimeter (clamps omitted).

In order that the frequency shall depend but little upon the elastic constants of the wire $2b\kappa/al$ must be small compared with 1. Now $b/a = \sqrt{[(q+T)/T]}$, and T cannot be increased indefinitely, or the wire will creep, and eventually will break. A safe limit is about $T/q = 1/2{,}000$.

κ may be made small by using a thin flat wire since $\kappa = d/2\sqrt{3}$ where d is the thickness of the wire.

In the practical case, we have $q = 1{\cdot}27 \times 10^{12}$ dynes/cm^2, $T\omega = 65 \times 981$ dynes, $\omega = 125 \times 10^{-6}$ cm^2, $d = 5 \times 10^{-3}$ cm.; then $\kappa = 1{\cdot}44 \times 10^{-3}$ cm., $l = 5{\cdot}2$ cm. Hence $T = 65 \times 981/125 \times 10^6$ dynes/cm^2, and $\sqrt{[(q+T)/T]} = 50$, giving $2b\kappa/al = 2{\cdot}8 \times 10^{-2}$. Thus the rigidity correction is only 2·8% of the total frequency, and the formula may legitimately be used.

It is clear from the above that the wire should be flat, and owing to the ease with which it may be hard-soldered, beryllium copper was used. The available suitable strip was 0·010 in. × 0·002 in. in cross section: using this size with a string approximately 5 cm. long, and a weight of approximately 65 gm., a resonant frequency of nearly 1,000 c/s. was obtained. The wire hangs vertically between the poles of a permanent magnet, the lines of force being horizontal and parallel to the plane of the strip. If the wire vibrates across the lines of force, an alternating E.M.F. is generated between its ends: conversely, if an alternating E.M.F. is applied, the amplitude and phase of the vibrations, and hence the current in the wire, depend on the frequency. The wire accordingly behaves as a resonant circuit element, but of rather low dynamic impedance. The wire is attached at the top to a block, which is adjustable for tilt and is insulated from the main body of the instrument; one connection is made to this block, and the weight is connected to another terminal by means of a fine beryllium copper hair-spring.

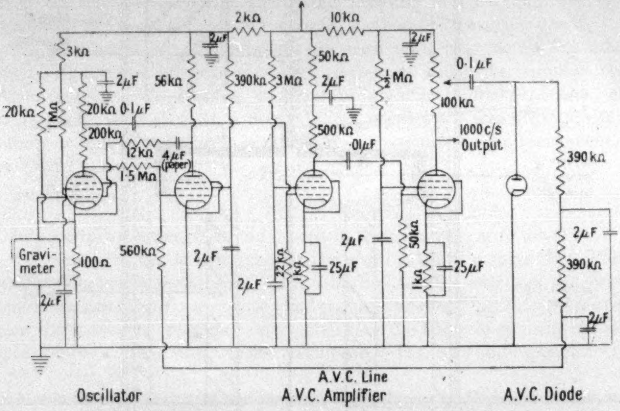

+200 V.
2 kΩ
10 kΩ
2μF
2μF
3kΩ
50kΩ
0·1μF
2μF
56kΩ
390kΩ
3MΩ
2μF
½MΩ
100 kΩ
20kΩ
1MΩ
20kΩ
0·1μF
200kΩ
500 kΩ
1000 c/s Output
·01μF
12kΩ
4μF (paper)
1·5 MΩ
390 kΩ
Gravi-meter
100Ω
560kΩ
2μF
2μF
22kΩ
1kΩ
25μF
50kΩ
1kΩ
25μF
2μF
390 kΩ
2μF
2μF
2μF
A.V.C. Line
Oscillator
A.V.C. Amplifier
A.V.C. Diode

Figure 2. Circuit diagram of driving circuit for gravimeter.

The circuit of the oscillator used for driving the string is shown in Figure 2. The voltage appearing across the string is amplified, and the amplified voltage fed back to the string. In order that the circuit shall act as an oscillator, the energy fed into the wire must be equal to the energy lost: owing to the low dynamic impedance of the string, this renders a two-valve amplifier necessary. The amplifier will "pull" the frequency of the string slightly from its natural value; the Q of the string is of the order of 20,000, and the phase change through the amplifier is about 1°. Thus the total pulling of the frequency is only 1 in 10^6, and variations in this will be less than 1 in 10^7. The pulling of the frequency makes it undesirable to use a transformer in the circuit, for if any direct current in the transformer should vary, the inductance will alter, and with it the phase change. Thus an extraneous frequency change would be observed.

The output frequency was compared with a crystal frequency standard. The method of comparison was suggested by members of the G.P.O. Research Laboratory, Dollis Hill. The thousand cycle signal from the crystal clock is split into three phases by condensers and resistances. Each phase is made to produce beats with the thousand cycles from the oscillator, thus producing three phase beats, which are applied to a small "Selsyn" motor. The motor makes one turn for each beat, and fractions of a beat may easily be estimated. The beats are counted by a gear train and dials. This method has the great advantage over most others, that it differentiates between frequencies greater and frequencies less than the comparison frequency: in the one case the Selsyn rotates clockwise, and in the other, anticlockwise. This is a very valuable feature for sea use, where vertical and horizontal accelerations may vary the apparent value of gravity by nearly one per cent. For, if the apparatus did not discriminate between frequencies greater or less than 1,000 c/s., then the nominal beat frequency would have to be greater than the maximum variations caused by the movement of the ship.

The integrating property of the instrument was pointed out above. If the apparent value of gravity in the submarine is fluctuating then the mean value measured, g, is to the first order

$$\frac{1}{T}\int_0^T g\,dt = g + \frac{(dx/dt)_T - (dx/dt)_0}{T}.$$

The first-order correction arising out of the difference of vertical velocity $(dx/dt)_T - (dx/dt)_0$ at the beginning and end of an observation could be eliminated if the dials were always photographed when $(dx/dt)_T - (dx/dt)_0 = 0$. This could be done by obtaining first a voltage proportional to d^2x/dt^2 and integrating: the camera could be arranged to trigger automatically whenever $(dx/dt)_T = (dx/dt)_0$. The technical difficulties of doing this were not solved in the present instance, however. The second-order correction remains in any case and is discussed below.

An electronic thermostat was constructed. An A.C. voltage was applied to a Wheatstone bridge, the arms of which consisted of coils, of different temperature coefficient but equal resistance at the balance point. The output from the bridge was amplified and applied to a phase sensitive rectifier, which discriminated between too high and too low a temperature, and controlled the heater accordingly. The coils were wound on a large copper can, two of 47 S.W.G. copper wire, and two of 32 S.W.G. constantan wire, thus providing four coils of the same number of turns, and with approximately the same resistance, but with temperature coefficients 0·0043° C. and zero respectively. The coils were wound similarly so as to balance out their self-inductance, and the bridge was wound so that there was no mutual inductance between the input and the output points. The heater was also wound non-inductively, to avoid stray magnetic fields. The 1,000 c/s. voltage from the crystal frequency standard was used for the A.C. supply, since it had a good waveform; a signal of 5 volts was applied to the bridge. The out-of-balance voltage for an increase in temperature of 0·001° C. from balance was 21 μv.; this was amplified by a factor of about 5×10^4, and when rectified, gave about 1 volt across a polarized relay, which was more than adequate to operate it. In practice, the hunting of the temperature was estimated at about 0·0005° c. on either side of the balance point.

The strip of beryllium copper was originally clamped at each end between blocks of brass, as in Figure 3, but no very accurate provision was made for ensuring that the edges of the blocks clamping the wire were level. This meant that, for half the cycle, the wire was effectively of a different length from that of the other half-cycle, which may be more clearly seen in the exaggerated sketch. Alternative methods of fixing were investigated. The next method used was to soft-solder the wire to two small half-cylinders of brass, with the wire lying along the axis of the cylinders. These half-cylinders then located the ends of the wire, and were clamped between the blocks, a "vee" groove being cut in one of the blocks to locate the half-cylinder. As the wire was soldered to the half-cylinders whilst held in a jig, this method ensured that the wire hung vertically, within the accuracy of machining. Experimental results encouraged us to proceed further to eliminate the zero drift, which was most pronounced when the weight was clamped and unclamped. This led to the possibility that when tension was applied suddenly on unclamping, the wire might slip slightly through the soft solder. Accordingly attempts were made to hard-solder the wire; this was finally accomplished in an atmosphere of nitrogen in an electric furnace. After soldering (in a jig) the wire was quenched, and then heat-treated to re-harden the beryllium copper strip, which had been softened by the high temperatures involved.

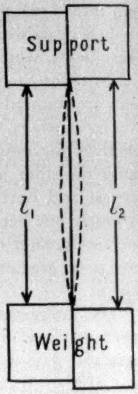


Figure 3.

The zero-drift with this string was indeed much less than that with the soft-soldered strings, but still was not negligible. This was felt to be due to the method of clamping, whereby the weight was lifted, and the tension removed from the string. It was therefore arranged to clamp the weight, yet keep the string under tension. In order to do this, some part of the assembly was mounted on a spring to prevent very slight movements of the weight when clamped from stretching the wire. This was accomplished by mounting the top end of the string on a spring. If suitably orientated, the spring will present a high mechanical impedance to the top of the wire, and yet allow vertical movement, which is the desired feature. Great difficulty was at first experienced in soldering the wire to the top end of the spring, owing to oxidation of the beryllium copper, even when in an atmosphere of nitrogen. A hydrogen furnace was used to solder the wire, but was later discarded in favour of a very small oxy-hydrogen flame. This enabled soldering to be effected without oxidizing the rest of the wire by overheating. The wire was subsequently heat-treated in the hydrogen furnace.

A clamping device was considered which could be operated without disturbing the apparatus. This consisted of a tube carrying two arms with rounded ends which press upwards on the weight when clamped. The tube slides in a vertical guiding tube, and as it moves upwards it operates two levers through a system of push-rods, each carrying a ball-bearing which moves downwards toward the weight. Consequently, the weight is clamped in the position in which it normally rests; indeed the frequency of the wire is very nearly unchanged when the weight is clamped. The motion of the clamps was produced by a "Sylphon" bellows, the top surface of which formed part of the wall of the vacuum system. When the under surface of the bellows was at atmospheric pressure, an upward force was produced on the bellows, and transmitted to the clamps. On evacuating the under compartment, the clamps were returned to their free position by springs.

Before the spring at the top of the wire was used, it was noticed that the un-sprung wire acted as a vertical seismograph. This was discovered by accidentally dropping a large weight on the floor. The cause was easily ascertained: for the tension in the string would change momentarily, although the mean tension over a time would remain constant. The frequency depends on the square root of the tension; therefore if the tension varies, the mean frequency will be lower than it would be under static conditions, by a correction of the second order of the variation of frequency. This is also the explanation of the second order correction due to vertical accelerations of a submarine at sea. The spring, however, will ensure that the tension in the wire remains more nearly constant during sudden vertical accelerations.

§ 3. LABORATORY TESTS

The effect of temperature on the frequency of the wire was found to be 0·0049 c/s. per ° c., in 1,000 c/s. This corresponds to approximately 10 milligals per ° c., over a range from 15° c. to 27° c., and is only about one third the coefficient expected from the known coefficient of expansion of the wire and the variation of its elastic properties with temperature.

The pressure effect appears to be due to three causes: firstly, the air has a buoyancy effect; secondly, a certain amount of air is carried with the wire, which increases the effective mass per unit length; these predominate at the higher pressures between 0·1 mm. and 760 mm.; thirdly, by virtue of its damping properties, the air lowers the Q of the wire, and hence the frequency changes, due to the fact that there is a phase-change in the amplifier. This third cause becomes apparent only at low pressures (below 0·1 mm.) where the first two effects are small. The variation of frequency with pressure is shown in Figure 4. The dynamic impedance of the string increased steadily as the pressure was reduced, a graph of the impedance plotted against the logarithm of pressure being a straight line within the limits measured. The impedance rises from 24 ohms at atmospheric pressure, to 80 ohms at 2×10^{-3} mm.

The drift experienced when the instrument was in use may be seen from the following results, obtained when the instrument was mounted in H.M. Submarine *Talent*.

Fort Blockhouse	19th May	Beat frequency	2·8607 c/s.
,, ,,	21st May	,, ,,	2·862 c/s.
Portland	27th May	,, ,,	2·866 c/s.

A change in beat frequency of 0·001 c/s. is equivalent to -2 milligals: the value of gravity at Blockhouse is 2 milligals greater than that at Portland.

The amplitude effect appears to depend on the cube of the amplitude: it is easier to maintain the amplitude of oscillation of the wire constant than to apply a correction. This is effected by means of an automatic volume control system in the oscillator, which maintains the voltage across the string at a constant value. In practice, this voltage is about 5 microvolts, which corresponds to an amplitude of about 10^{-5} cm. The noise level is of the order of a tenth of a microvolt: but the amplitude of oscillation was not reduced further, as the automatic volume control was sufficiently good to keep the error due to this cause below half a milligal. (The frequency is actually 0·005 c/s. lower than the frequency at zero amplitude when the apparatus is operating normally.) When the mains voltage was varied from 160 to 195, the frequency increased by less than 0·0005 c/s. in 1,000 c/s. Tilting the apparatus appeared to vary the frequency approximately as $1/\cos\theta$ if θ was small, a tilt of 1° from the centre frequency minimum causing an increase of frequency of 0·15 c/s. in 1,000 c/s. So long as the levelling is correct to within 0·05°, error due to this is less than 1 milligal.

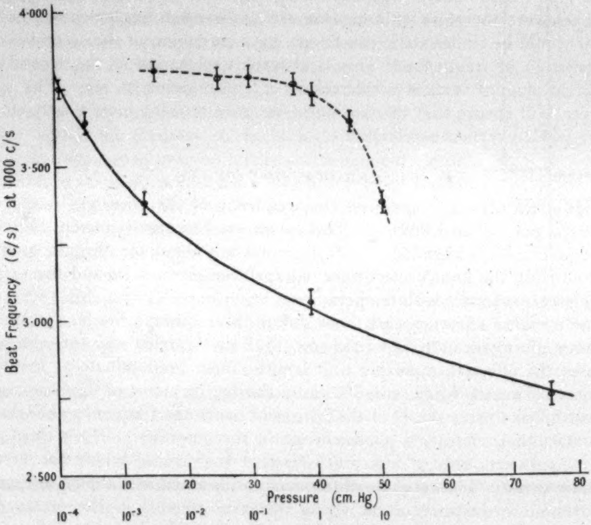


Figure 4. Effect of air pressure on frequency. The lower figures refer to the broken curve.

§ 4. FIELD TESTS

So far, the only field tests made on this instrument have been at sea during the recent cruise of H.M. Submarine *Talent* in the English Channel. A Vening-Meinesz 3-pendulum apparatus was used as a standard of reference. The instrument was mounted in gimbals to eliminate the effect of the slight rolling of the vessel even when submerged. Records were made automatically at 5-second intervals with a 16 mm. camera, which photographed together the beat-counter and a synchronous clock governed by the crystal frequency standard.

Several difficulties were encountered: electrical interference made it difficult to keep the string in oscillation: the apparatus accidentally became filled with oil while the pump was stopped, and the wire was damaged while adjustments were being carried out and had to be replaced. Despite these accidents, several readings were obtained which proved to be in very reasonable agreement with the pendulum apparatus.

The errors involved whilst at sea were mainly due to two causes: (*a*) imperfections in the gimbal system, which caused the gravimeter to be tilted by the rolling of the submarine, and (*b*) the movement imparted to the submarine by waves, which may vary the apparent value of gravity by one per cent with a period of about six seconds. The maximum estimated error due to the first cause is 3 milligals, and that due to the second, 1 milligal. The probable error of a pendulum observation is 1·5 milligal. The results obtained are shown in the table below.

Station number	Value of gravity [pendulums]	Reading of gravimeter	Error [mgals.]	Station number	Value of gravity [pendulums]	Reading of gravimeter	Error [mgals.]
84	981·080	981·076	−4	Cherbourg	981·043	981·042	−1
85	981·064	981·066	+2	93	981·060	981·055	−5
86	$981{\cdot}073_5$	981·072	−1·5	[94]	$981{\cdot}086_5$	981·080	[−6·5]
87	981·043	981·040	−3	95	981·085	981·090	+5
88	981·034	981·031	−3	96	981·076	981·078	+2
89	981·088	981·090	+2	97	981·110	981·106	−4
90	981·112	981·112	0	52*h*	981·105	981·108	+3
91	981·106	981·102	−4	52*i*	981·104	981·108	+4
92	981·078	981·080	+2	[52*j*]	981·098	981·106	[+8]

Root mean square value of errors = 4.

The two results given in brackets were taken under unusually rough conditions; if these are discarded the r.m.s. error is 3·2 milligals. The standard error of the pendulums is 2·4 milligals, giving a standard error of the same order for the gravimeter, and a probable error of 1·5 milligals for the gravimeter alone.

§ 5. CONCLUSION

Although only in the initial stages of development, the apparatus already shows great promise. There is still a small drift of frequency with time, but there are indications that this grows smaller as the wire ages. The results obtained at sea during the cruise in H.M. Submarine *Talent* show agreement between the pendulums and the gravimeter within the limits of experimental error of both. This is a great step forward as hitherto the use of pendulums has been the only successful method of measuring gravity at sea. The performance could be considerably improved by using better gimbals, and readings simplified by the use of further electronic aids.

If the drift can be reduced sufficiently, the apparatus will be of geodetic use (i.e. large variations of gravity may be accurately measured), and may even prove more accurate than pendulums. Investigations on the densities of rocks, by lowering the instrument down a borehole, should prove practicable. In order that this could be done, the apparatus would have to be permanently evacuated, the clamping mechanism changed so as to be operated from the surface, and the gimbals improved. The problem of thermal control of the instrument might be overcome by good thermal insulation and a recording thermometer. Owing to the great precision with which the frequency may be measured, the apparatus may have uses for investigation of earth tides at suitable permanent stations. The possibility of using it as a seismograph has to be looked into further.

ACKNOWLEDGMENTS

This work was begun by Mr. D. Boyd, of the Department of Geology, Glasgow University; his work during his stay in this Department did much towards indicating the difficulties which had to be overcome before the instrument reached its present stage.

Acknowledgment has to be made to the Department of Scientific and Industrial research, for the maintenance grant which makes this work possible, and to Mr. B. C. Browne, the author's supervisor, for many helpful remarks and suggestions.

REFERENCES

BERTRAND, G., 1938, *C.R. Acad. Sci., Paris,* **207,** 356.
RAYLEIGH, Lord, 1929, *Theory of Sound,* Vol. I (London: Macmillan), p. 296 ff.

Viscosity and Density in the Supercooled Liquid State.

BY C. DODD AND HU PAK MI
University College, London

Communicated by E. N. da C. Andrade ; MS. received 30th March 1949
read 4th February 1949

ABSTRACT. In order to determine whether any structural change takes place when a liquid is supercooled, measurements have been made on the viscosity of pure phenyl ether, which by its melting point (26·85° c.) is suitable for accurate work. It has been found that in both the ordinary and the supercooled state the liquid obeys a formula of the type $\eta = A \exp(b/T)$, but that the value of b in the supercooled state, which was investigated down to 17° c. below the melting point, is significantly larger than the value of b for liquid above the melting point. It is shown that an accurate value for the melting point can be obtained from viscosity measurements without the liquid ever becoming solid. The temperature coefficient of density shows no detectable change in passing through the melting point.

§ 1. INTRODUCTION

THE phenomenon of supercooling is well known. A substance can often exist in the liquid state even though its temperature may be several degrees below the melting point of the substance, and even though the liquid be stirred or made to flow. The liquid may remain supercooled indefinitely provided no solid crystal of the substance is dropped into it. The melting point of the substance may, however, be quite definite, as shown for example by the cooling curve for phenyl ether (Figure 1) taken with frequent additions of tiny crystals of the substance. The horizontal part of the curve gives the melting point at 26·85° c. It is of interest to know if any change in the nature of the liquid occurs as it is cooled and passes through the melting point into the supercooled region. A very suitable property for investigation of any such change in the liquid is that of viscosity, for on theoretical grounds a pure, non-associated liquid obeys the law given by Andrade (1934), viz. $\eta = A \exp(E/\boldsymbol{k}T)$, where η is the viscosity of the liquid at absolute temperature T, $\boldsymbol{k}$ is Boltzmann's constant, and A and E are constants for the particular liquid, whilst experimentally, relative viscosities may be measured to a very high degree of accuracy of the order of one part in 100,000.

§ 2. APPARATUS

A pipette viscometer was used, in which a fixed volume of the liquid is drawn up into the pipette and allowed to discharge into a reservoir (Figure 2). Any change in working volume by evaporation is avoided by the incorporation in the suction tube of a sphere S containing some of the liquid at the same temperature as that in the viscometer. Local distillation is prevented by immersing all glass parts having contact with the liquid or its vapour in a thermostatically controlled water bath, level LL. All ground joints are made water-resistant by the application of silicone grease. With a large, well-stirred water bath, the temperature can be maintained constant to within 0·005° c. which, for a time of flow of about 300 sec., is equivalent to a variation in flow time of 0·03 sec. Since the extreme deviation to be expected, using a stop watch to measure the time of flow , is 0·3 sec., this method was discarded in favour of a photoelectric method. The liquid is drawn up above the lower bulb and by adjustment of the tap is allowed to fall under its own hydrostatic head. An image of an illuminated slit is formed at A and the light passes on through the liquid and is focused on to a photoelectric cell controlling a high speed relay through a suitable amplifier. As the liquid meniscus passes A it scatters the light momentarily and the cell operates the relay, which starts an electric clock. A similar interruption of a beam of light when the meniscus reaches B is used to stop the clock, which is of the synchronous motor type controlled by a tuning fork, and has four dials reading down to 0·01 sec. The mean deviation for ten successive times of flow never exceeded 0·03 sec., showing that full use was being made of the constancy of the temperature

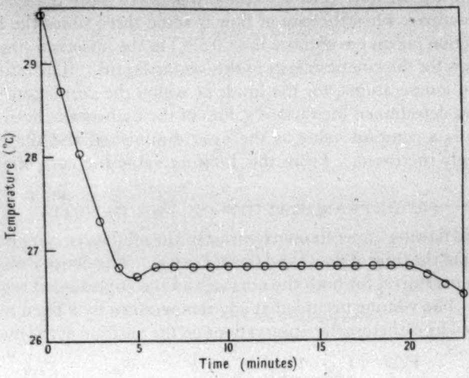


Figure 1.

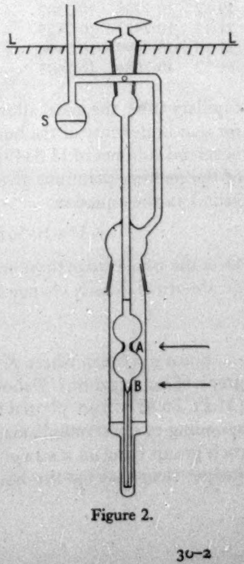


Figure 2.

provided by the thermostat. Over a limited range of temperature the working volume changes only slightly, due to thermal expansion; therefore, rather than attempt to keep the working volume constant, the change in time of flow at a given temperature, produced by adding small volumes from a 1 cm^3 pipette, was found, and an appropriate correction deduced from the known coefficient of expansion of the liquid. The correction was never more than 0·5%. At the higher temperatures, when the rate of flow is some three times the lowest value, a small correction (again never more than 0·5%) to the observed time of flow was applied to allow for the kinetic energy of the issuing liquid. The ratio of the times of flow at two temperatures, for the lower of which the correction was certainly negligible, was determined for various values of the hydrostatic head. This ratio tended towards a constant value as the head diminished and the times of flow correspondingly increased. From this limiting value the correction was determined.

§ 3. DETERMINATION OF THE DENSITY

For a liquid flowing under its own hydrostatic head, the viscosity is proportional to the product of the time of flow t and the density ρ. The density was determined at various temperatures, for both the normal and the supercooled regions, using a dilatometer. The volume occupied at any temperature by a fixed mass of liquid was determined by cathetometer observations on the position of the meniscus in the capillary tube, the usual allowance being made for the expansion of the vessel and for non-uniformity of the bore of the calibrated capillary tube. In the Table the observed volumes of 11·5349 gm. of phenyl ether at temperatures θ° c. on each side of the melting point are given, together with the error obtained by fitting these values to the equation

θ (° c.)	V_{obs} (cm^3)	V_{calc} (cm^3)	$\frac{V_{\text{obs}}-V_{\text{calc}}}{V}$	θ (° c.)	V_{obs} (cm^3)	V_{calc} (cm^3)	$\frac{V_{\text{obs}}-V_{\text{calc}}}{V}$
12·40	10·6667	10·6664	$+2{\cdot}8\times10^{-5}$	25·98	10·7846	10·7849	$-2{\cdot}8\times10^{-5}$
13·75	10·6782	10·6780	$+1{\cdot}9\times10^{-5}$	26·85	Melting Point		
14·07	10·6811	10·6808	$+2{\cdot}8\times10^{-5}$	28·01	10·8025	10·8029	$-3{\cdot}7\times10^{-5}$
16·04	10·6978	10·6978	$0{\cdot}0\times10^{-5}$	30·17	10·8224	10·8222	$+1{\cdot}9\times10^{-5}$
18·13	10·7157	10·7159	$-1{\cdot}9\times10^{-5}$	32·24	10·8405	10·8407	$-1{\cdot}9\times10^{-5}$
19·82	10·7308	10·7307	$+0{\cdot}9\times10^{-5}$	33·82	10·8548	10·8551	$-2{\cdot}8\times10^{-5}$
21·17	10·7427	10·7425	$+1{\cdot}9\times10^{-5}$	35·27	10·8670	10·8679	$-8{\cdot}3\times10^{-5}$
21·98	10·7499	10·7496	$+2{\cdot}8\times10^{-5}$	36·02	10·8740	10·8748	$-7{\cdot}4\times10^{-5}$
24·12	10·7684	10·7685	$-0{\cdot}9\times10^{-5}$				

$$V = 10{\cdot}5612(1 + 7{\cdot}9188\times10^{-4}\,\theta + 9{\cdot}033\times10^{-7}\,\theta^2).$$

Over the whole range these errors are small and unsystematic, showing that there is no abnormal density change as the liquid passes through its melting point.

§ 4. VISCOSITIES

Since $\eta = K(t\rho)$, where K is an apparatus constant, then if $\eta = A\exp(E/\mathbf{k}T)$, a graph of $\log\eta$ against $1/T$ should give a straight line. The results for phenyl ether (M.Pt. 26·85° c.) are plotted in Figure 3. The points fall into two groups corresponding to observations made above and below the melting point, the points of each group lying on a straight line, but the line for the supercooled liquid being steeper than that for the liquid above the melting point as shown in Figure 4.

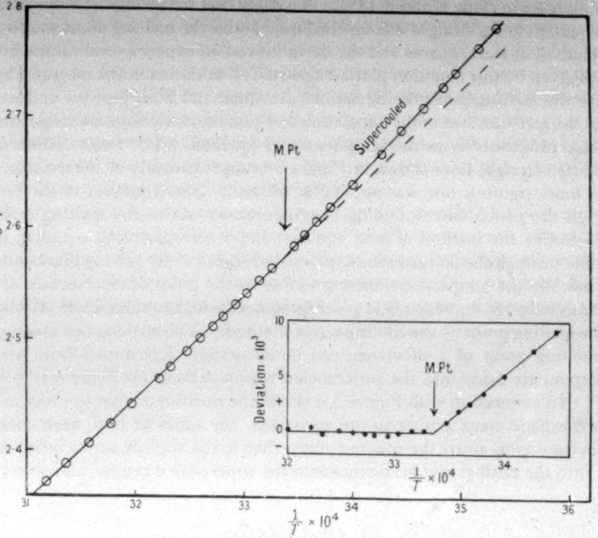


Figure 3.

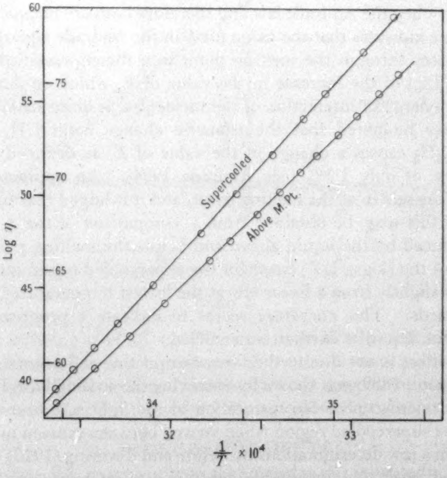


Figure 4. Graphs of log η against $1/T$ for phenyl ether above the melting point (outer scales) and supercooled (inner scales).

The increase in slope is about 13%. To show that the change of slope is abrupt, the equation to the straight line for the liquid above the melting point was found by the method of least squares and the deviations of *all* experimental values from the values given by this equation plotted against $1/T$ as shown in the inset to Figure 3. Above the melting point the deviations are small and both positive and negative about the zero line, but in the supercooled region the deviations are all positive and increase progressively as the temperature is lowered. The point of intersection of the two straight lines shown in Figure 3 is approximately at the melting point, but a more rigorous test was applied as follows. The equation to the best line through the points corresponding to temperatures above the melting point was calculated by the method of least squares, and a corresponding equation for the best line through the points in the supercooled region. By solving these equations algebraically the temperature corresponding to the point of intersection of these two lines is 26·95° C., which is in good agreement with the value 26·85° C. obtained for the melting point by the cooling-curve method. The striking fact emerges that the melting point of a substance can be accurately determined from viscosity measurements made into the supercooled region *without the liquid ever becoming solid*! In connection with Figure 3 it should be mentioned that in order to avoid any systematic error arising in the apparatus, the times of flow were measured first in the region above the melting point, then in the slightly supercooled region, back into the normal region, further into the supercooled region, and so on.

§ 5. DISCUSSION

The fact that the graph of $\log \eta$ against $1/T$ gives one straight line for the heated liquid and another for the supercooled liquid is interesting since it shows that both approximately obey the Andrade law and therefore both are normal liquids. The change in slope indicates that the value for E in the Andrade equation increases as the liquid passes through the melting point into the supercooled region. The magnitude (13%) of the increase in the value of E, which on Andrade's theory represents an energy of interaction of the molecules, is quite striking. For comparison it may be noted that the isomeric change from $CH_3—O—C_3H_7$ to $C_2H_5—O—C_2H_5$ causes a change in the value of E, as deduced from viscosity measurements, of only 1·3% (see Andrade 1934). An appreciable structural change therefore sets in at the melting point, and it is hoped that further evidence in support of this may be obtained from a comparison of the X-ray diffraction patterns produced by the liquid above and below the melting point. A careful examination of the ($\log \eta$, $1/T$) curve for the supercooled liquid indicates that the curve departs slightly from a linear law at the lowest temperatures, being slightly concave upwards. This curvature seems to indicate a progressive change in structure as the liquid is further supercooled.

That the effect is not due to the formation of the well-known liquid crystals (e.g. see Lehmann 1889) was shown by observing the cooling liquid with polarized light under a microscope. No restoration of the light was found as the liquid passed into the supercooled region when viewed between crossed nicols.

Apart from a few determinations by White and Twining (1913) of the viscosity of supercooled water (an unsuitable substance as its temperature variation of viscosity differs markedly in form from that of the vast majority of liquids), the only systematic determination on a supercooled liquid has been made on liquid white phosphorus by Dobinski (1934). Although the accuracy of these determinations is not great (about 2%), he obtains a curve similar to Figure 3, but his results indicate that the departure from the linear law occurs some 5° c. above the melting point, which conflicts with our results. When it is borne in mind that phosphorus can exist in various allotropic forms and that on supercooling the clear liquid phosphorus becomes turgid and opaque (a phenomenon which is absent in phenyl ether), it is clear that the mechanism of the viscosity anomaly in phosphorus is different from that for phenyl ether discussed here.

The viscosity measurements are being extended to other liquids, including liquid gallium, which readily supercools and whose structure is comparatively simple compared with diphenyl ether, which was chosen initially purely because of its convenient melting point. Curves for log η against $1/T$ are plotted in Figure 5 for m-chlor-nitrobenzene (melting point 44·2° c.), the point of intersection of the two straight lines being at 44·28° c. The deviations of all points from the straight line through points taken at temperatures above the melting point are plotted in the insert to Figure 5. It is hoped also to carry out determinations of the surface tension and of the refractive index into the supercooled region.

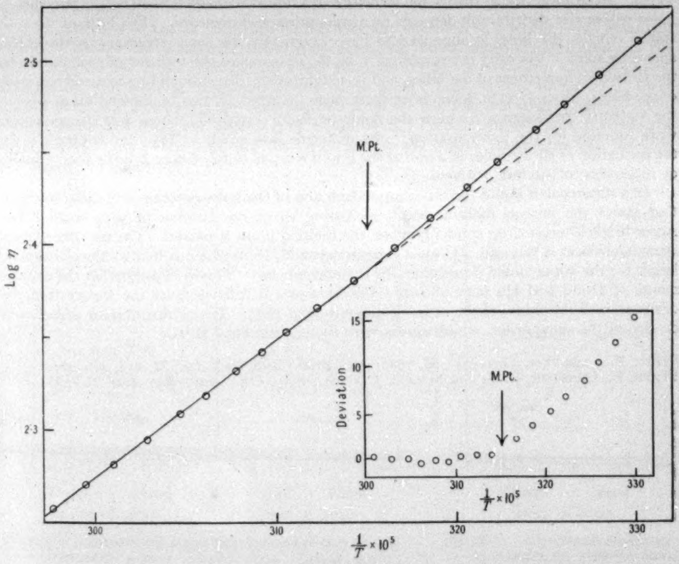


Figure 5.

ACKNOWLEDGMENTS

We wish to take this opportunity to acknowledge our indebtedness to Professor E. N. da C. Andrade for suggesting the original problem and for the considerable

assistance and encouragement which he has afforded us during this work. We also thank the Inspector of Naval Ordnance, Woolwich, for the loan of the phonic electric clock and associated equipment.

REFERENCES

ANDRADE, E. N. DA C., 1934, *Phil. Mag.*, **17**, 698.
DOBINSKI, S., 1934, *Bull. int. Acad. Cracovie*, 3–4 A, 103.
LEHMANN, O., 1889, *Z. phys. Chem.*, **4**, 462.
WHITE, G. F., and TWINING, R. H., 1913, *Amer. Chem. J.*, **50**, 380.

DISCUSSION

Dr. R. FÜRTH. I think that the most interesting results of Dodd and Hu can be explained on the basis of the " hole " theory of the liquid state (Fürth *et al.* 1939, Fürth 1940, 1941). According to that theory the structure of a liquid is characterized by the presence of holes whose size distribution depends on temperature and pressure. The bulk of the substance outside the holes is supposed to have practically the same structure as the corresponding solid. Viscosity is considered to be the outcome of the transfer of momentum by the Brownian movement of the holes, and its magnitude is determined by the mean free path of the holes, which in turn depends on their mean lifetime. It can be shown that as a result the viscosity coefficient η contains the Andrade factor $\exp \phi/\boldsymbol{k}T$, where ϕ is the molecular work function for the " evaporation " of a molecule into a hole. The total energy ΔU for the formation of all the holes in a mol of the liquid is equal to ϕZ, where Z is the total number of molecules at the hole surfaces.

In a supercooled liquid the distribution function of the holes is essentially different from that above the normal melting point, indicating an excess number of very small holes. Accordingly Z must drop suddenly when the melting point is passed. On the other hand, as no latent heat is released, ΔU must remain constant, from which it follows that ϕ must be larger for the supercooled liquid than for the normal one. This is exactly what the experiments of Dodd and Hu have shown. Furthermore, it follows from the theory that the decrease in Z should be of the order of magnitude of 10%. Hence the increase in ϕ should be also of the same order, which agrees with the experimental facts.

FÜRTH, R., 1940, *Proc. Phys. Soc.*, **52**, 768; 1941, *Proc. Camb. Phil. Soc.*, **37**, 252, 276, 281.
FÜRTH, R., ORNSTEIN, L. S., and MILATZ, J. M. W., 1939, *Proc. Amsterdam Acad. Sci.*, **42**, 107.

LETTERS TO THE EDITOR

The Radiation Pattern of Retinal Receptors

I have recently made use of microwave theory in order to explain the directional sensitivity of the retina (Toraldo di Francia 1948, 1949). As regards the propagation of light in their interior, retinal receptors must be thought of as dielectric waveguides, while the more general problem of reception must be dealt with by means of the theory of antennas. Each elementary receptor is a microantenna, and its radiation pattern, which, on account of the reciprocity theorem (Schelkunoff 1943, p. 478), is the same in emission as in reception, is responsible for the directivity known as the Stiles and Crawford effect. I pointed out that the absence of the effect for the rods could be accounted for by their diameter being smaller and their radiation pattern wider than those of the cones. If the principal lobe of the rod pattern had a larger angular aperture than the pupil, the directivity of the rods could have remained concealed.

I am indebted to Dr. W. S. Stiles for drawing my attention to a recent paper (Flamant and Stiles 1948) where some evidence is given regarding the existence of a directivity of the rods. Making use of the fact that for a test stimulus of short wavelength, which is viewed parafoveally, the absolute threshold of the cones greatly exceeds that of the rods, these authors were able to separate the cone diagram from the rod diagram with certainty. The results obtained by these two observers (F.F. and W.S.S.) are shown in Figure 1, where the upper curves relate to the rods and the lower ones to the cones. It appears that the rods have almost no directivity.

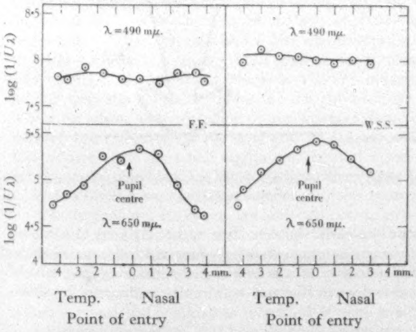


Figure 1. Directional sensitivity curves for two normal observers. Upper curve: rods. Lower curve: cones. (From Flamant and Stiles 1948, by kind permission of the authors.)

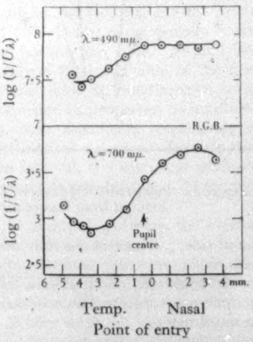


Figure 2. Directional sensitivity curves for an observer having the maximum sensitivity point displaced from the pupil centre. (From Flamant and Stiles 1948, by kind permission of the authors.)

However, the same authors have published the curves obtained with the observer R.G.B. (Figure 2), who presents in a high degree the not rare phenomenon (Stiles 1939, Flamant 1946, Le Grand 1948) of the point of maximum sensitivity being displaced from the pupil centre. This is a lucky circumstance for the researcher, since it allows the radiation pattern to be measured up to a wider angle than usual. In fact the upper curve of Figure 2 gives evidence in favour of the hypothesis that the rods too have a directivity which in ordinary cases is masked by the remarkable angular width of the major lobe.

It seems to me that Figure 2 shows another extremely interesting fact. Almost every antenna pattern of high directivity presents, besides the major lobe, one or more side lobes of minor importance, and I had been wondering for some months whether the side lobes of the cones could be revealed by a suitable experiment. The answer to this question is given again by Figure 2. From the data of the lower curve I have been able to plot the sensitivity of the cones against the angle of incidence with respect to their axis. The polar diagram thus obtained is shown in Figure 3, where the curve has been completed by symmetry on the nasal side. A part of the first side lobe is clearly visible. It is rather striking to compare this radiation pattern with that of a dielectric antenna, shown in Figure 4.

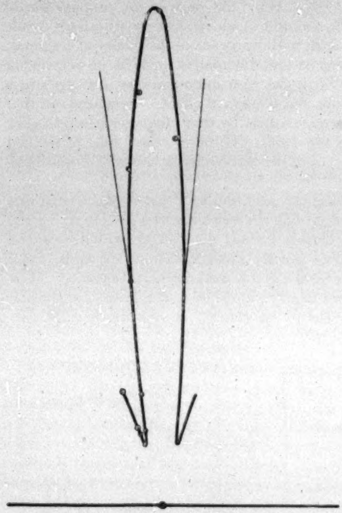


Figure 3. Polar radiation pattern of the cones, deduced from Figure 2.

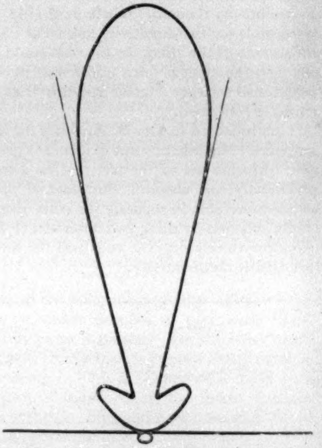


Figure 4. Polar radiation pattern of a dielectric antenna radiating from one of its ends.

It would seem advisable in the future to give all the directional sensitivity diagrams of retinal receptors in polar coordinates, as is done in Figure 3 and in all standard textbooks on antenna theory.

Institute of Physics of the University,
Firenze, Italy.
16th April 1949.

G. Toraldo di Francia.

Flamant, F., 1946, *Comm. Lab. Inst. Opt.*, **2,** 47.
Flamant, F., and Stiles, W. S., 1948, *J. Physiol.*, **107,** 187.
Le Grand, Y., 1948, *Rev. d'Opt.*, **27,** 759.
Schelkunoff, S. A., 1943, *Electromagnetic Waves* (London and New York : McGraw-Hill).
Stiles, W. S., 1939, *Proc. Roy. Soc.* A, **127,** 64.
Toraldo di Francia, G., 1948, *Nuovo Cimento*, **5,** 589 ; 1949, *J. Opt. Soc. Amer.*, **39,** 324 ; *Atti Fond. G. Ronchi*, **3,** 137.

REVIEWS OF BOOKS

Chambers' Six-figure Mathematical Tables, by L. J. COMRIE. Volume 1, Logarithmic Values, pp. xxii+576 ; Volume 2, Natural Values, pp. xxxvi +576. (Edinburgh and London: W. & R. Chambers Ltd., 1948.) 42*s*. each volume.

Chambers' *Mathematical Tables* is probably the best known set of tables (with the possible exception of the Board of Education's four-figure set), and the present two volumes are in some sense the successor to it, for it appears that in 1944, the centenary of their publication, Dr. Comrie was asked to report on that work with a view to bringing it thoroughly abreast of modern requirements, and that the new work is the outcome.

The bulk of the old book was a 200-page table of seven-figure logarithms and just over 130 pages of natural and logarithmic trigonometric functions; the remainder of its 454 pages were occupied with a miscellany of geodetic and navigational tables. The new work, even if a successor to the old, is in no sense a revision of it. Dr. Comrie has considered first the requirements of the present-day physicist and statistician and provided the tables he is likely to need, to an accuracy which seems adequate for most work. The logarithms are now to six figures, and occupy 180 pages, but data for obtaining eight-figure logarithms are also provided (the old book gave ten figures for argument 1 to 1·06 at intervals of 0·0005).

In volume 2, the more interesting of the two, we have, for $x=1$ to, 1000, the square, cube and fourth power (fifth power up to 100), the reciprocal, the factorial and the prime factors as well as $\log x!$ and $x^{\frac{1}{2}}$, $(10x)^{\frac{1}{2}}$, $x^{\frac{1}{3}}$, $(10x)^{\frac{1}{3}}$ and $(100x)^{\frac{1}{3}}$ and some powers of the reciprocals. The factorials above 200 are here published for the first time. From 1,000 to 3,400 we have the square, cube and prime factors. There is also a table of the first 1,540 primes, the largest of which is 12,919; evidently this table is in no sense a competitor with the specialist tables for the use of those interested in properties of numbers. There are three sets of tables of the circular functions, with arguments respectively in degrees and minutes, in radians, and in degrees and decimals of a degree. In each case the cot and cosec for small angles are given at a reduced interval, and the usual auxiliary functions are provided for the functions where they become excessively large. There is a table for conversion from rectangular to polar coordinates.

The hyperbolic functions are well tabulated, as well as natural logarithms, the inverse hyperbolic (and circular) functions and the Gudermannian with its inverse.

Beyond these elementary functions, a selection of transcendents has been made, the choice falling on the gamma function and the probability integral in its two main forms. Here the range has been most carefully considered so that the danger of a user wanting values beyond the arguments tabulated is very small. Where so much is provided it is ungenerous to complain, but I think many a physicist would have welcomed some tabulation of elliptic functions, if only in the form of the elliptic integrals E and F.

This very short account of the volume, which does not do justice to the care taken in the arrangement of the tables to make interpolation easy and to ensure the maximum convenience in use, also fails to mention another very valuable feature—the little treatise on numerical differentiation and differential equations, which comes near the end. There is also a great deal of really valuable information in the introduction, and a very helpful bibliography of larger tables at the end.

In volume two, the 6- and 8-figure logarithms have been mentioned. There is also a table of antilogarithms, a function for which tables to more than four figures have been remarkably scarce up to the present. Naturally, too, there are tables of the logarithms of trigonometric functions, which, as in volume 1, are given for angles expressed in degrees and minutes, or in degrees and decimals, or in radians. The S and T functions are also given for all these arguments and for seconds of time as well. There are tables of $\log \Gamma(x)$ and of the logarithms of the hyperbolic functions.

Taking the two volumes together, it seems reasonable to suppose that they will suffice for the everyday needs of most workers who have calculations to do, though each will supplement them with tables for his own special work; perhaps he will want a selection of tables for statisticians, or tables of Bessel functions, or maybe of factors. From this point of view the collection is not the successor to Chambers' Tables, but to such collections as those of Hoüel or Potin, Barlow, Hutton, Dale or Dwight. Certainly, with Barlow's tables on one side of the new book and Jahnke and Emde on the other, any worker would have a fine collection of tables at hand.

From a footnote to the preface it appears that there is also an abridged edition of these 6-figure tables, in one volume, but the exact contents are not known to the reviewer. There is also a suggestion that the 64-page book of four-figure tables prepared by the same compiler for Messrs. Chambers is a miniature version of the present two volumes. It is certainly a very handy volume, but to the present writer it would have seemed that a nearer analogue is *Standard Mathematical Tables* by Milne-Thomson and Comrie.

J. H. A.

Some Aspects of Oculo-Refractive Technique, by MALCOLM CHOLERTON. Pp. 148. (London: Hammond, Hammond & Co. Ltd., 1948.) 21*s*. net.

Since numerous physical, physiological and psychological factors are involved in even the simplest visual act in any visual experiment or test it is extremely difficult for the observer to rid himself of the psychological elements of experience and memory and to extricate the simple sensation from the complex mental precept which he experiences. Snow is always " white " and coal is " black " whatever may be the intensity and spectral composition of the light by which they may be illuminated. When experimenting on vision we are dealing with the consequences of the interaction between a physical agency, radiation and living organisms ; and running parallel with these physiological actions, consciousness is involved. The analytical methods of physical science, in which the subject of enquiry can be resolved into elementary concepts such as atoms and electrons, cannot be applied or can be applied only in a modified form.

Because of the confusion of all these factors, progress in visual science has been much less impressive than in physical science ; and its literature is frequently misleading. And how difficult it is to define a " normal observer " and to know whether such a one is operating! Not only may an observer classified as normal be temporarily upset physiologically, but he may, perhaps for a reason over which he has no control, be incapable of the normal interest in or attention to the work in hand, or of eliminating some psychological factor that is vitiating the results.

These difficulties may be present during clinical testing by the optical practitioner. Fortunately, however, such is the adaptability of the eye and visual mechanism, they do not obtrude unduly in the majority of cases. The physical tests which the practitioner applies, with such instruments as the ophthalmoscope to search for possible signs of disease, and with lenses and physical appliances to measure and treat ametropia and oculo-motor imbalance, usually provide acceptable comfort.

To meet those cases in which the orthodox clinical methods prove to be inadequate, various methods are practised by enterprising practitioners seeking to improve their service to the community, but they are empirical methods and often unsuccessful. More rational methods are not yet possible because our knowledge of the visual process and of the nature of vision, particularly binocular vision, is too scanty.

There is much discussion of practical clinical problems amongst opticians, but in the absence of the needed basic knowledge of visual science such discussion amounts to little more than dialectics in which the terms used are not precisely defined ; and little or no progress is made. Investigation is sorely needed by those who are skilled in the methods of research and who will pay heed to the biological aspects of the problems.

The book under review is concerned with certain clinical methods which will be of interest to established optical practitioners. It does not attempt a really critical review of such methods but sets out, to quote from the author's preface, " to bring together, and discuss generally, the various phases of the clinical examination of the eye and visual function in so far as they extend and should concern the optician ". Chapters are devoted to general routine, anisometropia and aniseikonia, presbyopia, dynamic retinoscopy, external and internal examination, visual fields, colour vision testing, tinted lenses, orthoptics, secondary correspondences and lighting practice.

Under several of these headings, however, it is not discussion that is needed, but experiments and yet more experiments. The author of this book draws attention to the prevailing confusion of terminology, particularly concerning dynamic retinoscopy, but unfortunately his own style of presentation is, to this reviewer at least, frequently obscure.

For example, in the preface the author expresses his dissatisfaction " with the efficacy of our present clinical approach to the analysis and rehabilitation of those forms of visual affliction or implication not wholly amenable to solution by the present means at our disposal ". He claims that opticians " are skilled in the physics of the situation, in the differential weighting of ametropia and anomalies of binocular vision in terms of the relative symptom-potential factor . . . . . ". In dealing with the fusional reserves in the chapter on general routine, and having referred to the blur or break point and negative vergence reflex phase, he says, " this involves consideration of the aspect whether a blur point can be demonstrated by the survival of the negative vergence reflex to a stage which can impose upon the accommodation a preceptual disjugacy with the distance of the target, by extending it reflexly further in association with the still surviving negative vergence reflex, until the latter is finally overwhelmed."

Although this style, which is general throughout the book, is difficult to follow, practising opticians will find much to stimulate them. H. H. E.

A Textbook of Radar, by the Staff of the Radiophysics Laboratory, C.S.I.R., Australia. Pp. vi + 579. (London: Chapman and Hall, Ltd., 1948.) 50*s*.

The relaxation of military secrecy since the end of the last war has brought forth a spate of books on radar and related subjects. The present book is the work of the wartime radiophysics team in Australia. It was first published in Australia and it is now being made available to British readers through Messrs. Chapman and Hall, Ltd. Like the earlier American and Italian books, it shows evidence of having been written in the latter half of 1945. For instance no references are given to the important papers on radar published in the *Journal of the Institution of Electrical Engineers* during 1946 or the corresponding American publications. This is a pity for a book published in 1948, and it will limit the value of the book for the serious student of radio.

There are twenty chapters in the book and twenty-one contributors. The editing has been carefully done, however, and in consequence there are few obscurities and few definite mistakes. The first two chapters deal with the fundamentals of the subject and cover the usual subjects—special features of radar, components of a radar set, the echoing process, the free space radar equation, normal propagation, super refraction, atmospheric absorption, and atmospheric scattering. The explanations are clear and concise and should be easily understood by anyone with a pre-war knowledge of radio.

Chapters III, IV and V deal respectively with the magnetron, triode power oscillators, and modulators. Both the theoretical and the practical aspects of these devices are discussed and examples of design problems are included.

The following four chapters continue the story on radar components and deal with transmission lines, waveguides, aerials and aerial duplexing. They occupy 210 pages, and a very full discussion of the basic principles and the design of these components is given. The examples mentioned cover a wide field. Thus, in the chapter on aerials, there are sections on lens aerials, metal plate lenses, electromagnetic horns, Foster scanners and Musa systems as well as the more orthodox types. There are, however, one or two omissions. Thus little is said about the important subject of back-to-front ratio, and as the aerial chapter includes a detailed discussion on the theory of linear arrays it would have been helpful to have corrected the widely believed impression that the power gain of aerial system is theoretically limited by its size. However, in spite of these omissions these four chapters will be of real value to the student.

The later chapters in the book (X–XV) deal respectively with the characteristics of the radar receiving system—the receiver, the timing unit and the presentation system. Separate chapters are included on local oscillators, frequency converters and amplifiers, and there are chapters on display circuits and automatic ranging circuits. For the most part the authors have concentrated on giving a dissertation on the principles of circuit design as applied to these components, but several examples of complete circuits are given with full explanations.

The final section of the book (Chapters XVI–XX) discusses radar systems. This is prefaced by an introductory chapter dealing with functional and operating characteristics of radar systems. Bearing in mind the length of the book, this chapter could have been rather longer with advantage. The later chapters deal with ground radar, shipborne radar, airborne radar and radar navigation respectively. All the main types of equipment are mentioned and military as well as civil uses are discussed. The treatment is concise and for the most part accurate. The authors have made one or two small factual errors in describing British radar equipment, but these are only of historical significance. An amusing mistake occurs on page 515, where the authors, in describing the CH system, state that the centres of both the upper and lower transmitting and receiving aerials were at the same height. Actually they weren't, but they should have been !

The book is well printed and bound and should be of value to any serious student of radio.

DENIS TAYLOR.

CONTENTS FOR SECTION A

PAGE
Mr. C. H. Miller, Mr. D. A. Long, Dr. L. A. Woodward and Dr. H. W. Thompson. A Recording Spectrometer for Raman Spectroscopy . . 401
Mr. L. L. Green and Mr. W. M. Gibson. The Neutrons Emitted in the Disintegration of Lithium by Deuterons 407
Prof. N. F. Mott. The Basis of the Electron Theory of Metals, with Special Reference to the Transition Metals 416
Dr. N. R. Campbell and Dr. L. Hartshorn. The Experimental Basis of Electromagnetism—Part III: The Magnetic Field 422
Dr. L. Hartshorn. The Experimental Basis of Electromagnetism—Part IV: Magnetic Materials 429
Mr. R. W. Wright and Dr. J. P. Andrews. Temperature Variation of the Electrical Properties of Nickel Oxide 446
Letters to the Editor:
  Mr. B. N. Watts. Increased Sensitivity of Infra-Red Photoconductive Receivers 456
  Dr. R. Wilson and Mr. G. R. Bishop. The Decay Constant of Radio-Sodium, ^{24}Na 457
  Mr. F. K. Goward, Dr. E. W. Titterton and Mr. J. J. Wilkins. The Photo-Disintegration of Oxygen into Four Alpha-Particles 460
Reviews of Books 461
Contents for Section B 463
Abstracts for Section B 463

ABSTRACTS FOR SECTION A

A Recording Spectrometer for Raman Spectroscopy, by C. H. Miller, D. A. Long, L. A. Woodward and H. W. Thompson.

ABSTRACT. A description is given of a photoelectric recording spectrometer for Raman spectroscopy incorporating a high-intensity excitation system and a novel amplification device. Operation of the exciting mercury arc lamps from 50 c/s. mains imparts to the light signal, and consequently to the current from the photo-multiplier tube, a 100 c/s. periodicity. After amplification by a narrow band-pass amplifier, rectification is achieved in a push-pull phase-sensitive (homodyne) rectifier, by " mixing " the signal with a fixed voltage of identical periodicity. The rectified signal is fed through a filter to a cathode follower valve which provides the power to operate a recording milliammeter. This method of rectification, in conjunction with appropriate filter, effectively reduces the bandwidth of the amplifier, with consequent great improvement in the signal/noise ratio. With this arrangement cooling of the photo-multiplier is unnecessary for normal work. Using a spectrometer of small aperture and resolution and a Raman tube of 35 ml. capacity, traces of the Raman spectrum of carbon tetrachloride show no appreciable background noise.

The Neutrons Emitted in the Disintegration of Lithium by Deuterons, by L. L. Green and W. M. Gibson.

ABSTRACT. The photographic plate method has been used in an investigation of the neutrons from the reaction

$$^7_3\text{Li}+{}^2_1\text{D}\rightarrow {}^8_4\text{Be}+{}^1_0\text{n}.$$

The results are in general agreement with those of Richards, but show the structure of the neutron groups more clearly. In particular, the shape of the broad group attributed by Richards to a level in ^{8}Be at $3{\cdot}0\pm0{\cdot}4$ MEV. is discussed in detail.

The total energy release is $15{\cdot}0\pm0{\cdot}15$ MEV., and evidence is put forward for the existence of energy levels in ^{8}Be at 2·8 MEV., 4·05 MEV., 4·9 MEV., and 7·5 MEV.

The Basis of the Electron Theory of Metals, with Special Reference to the Transition Metals, by N. F. MOTT.

ABSTRACT. It is shown that the collective electron and London–Heitler models are not to be regarded as different approximations to the same exact wave function for solids in which, according to the former model, there is a partially filled zone of energy levels. It can thus be shown why nickel oxide in the pure state is a non-conductor, although it contains an incomplete zone. The properties of the metals nickel, palladium and platinum are discussed in the light of these results ; platinum differs from nickel in that the orbital contribution to the moment of the elementary magnets is not quenched. A discussion is given of x-ray absorption edges, and it is shown why exciton lines are absent for metals.

The Experimental Basis of Electromagnetism—Part III: The Magnetic Field, by N. R. CAMPBELL and L. HARTSHORN.

ABSTRACT. The principles outlined in previous parts, published in the *Proceedings of the Physical Society* in 1946 and 1948 respectively, which dealt with the direct current circuit and with electrostatics, are here applied to magnetism, with the object of showing how the basic concepts are defined in terms of the operations actually performed in measuring them. This part is confined to a discussion of the vector **B**, which is shown to be measurable independently of a knowledge of any other magnitude. It is solenoidal and can therefore be determined at all points, even those within solid bodies.

The Experimental Basis of Electromagnetism—Part IV: Magnetic Materials, by L. HARTSHORN.

ABSTRACT. This paper concludes the inquiry which was undertaken by N. R. Campbell and the author with the object of elucidating the connection between the concepts and principles of electromagnetism and the experimental operations actually performed in the Laboratory, that constitute their factual basis. In Part III the vector **B** was established as measureable everywhere, even within solid bodies. The vector **H** and the scalar $\mu = \mathbf{B}/\mathbf{H}$ are now shown to be measurable in special circumstances by means of the magnetometer and permeameter, but in general their values depend on a hypothesis, which is stated. The significant facts concerning the magnetic properties of real materials are briefly reviewed, and the two-fold aspect of the science is emphasized: the self-consistent mathematical theories based on postulates, on the one hand, and the complex of experimental laws on the other. It is important that the relation between the two shall be clearly established, and not merely assumed as self-evident.

Temperature Variation of the Electrical Properties of Nickel Oxide, by R. W. WRIGHT and J. P. ANDREWS.

ABSTRACT. An experimental method of preparing specimens of NiO in the form of coherent strips is described. Simultaneous measurements of electrical conductivity σ and Hall effect R are then made at various temperatures up to about 700° c. and the thermo-electric power dE/dT measured immediately afterwards, all on one sample. The experiments are then repeated on different samples, and they show that log σ, log R, and dE/dT are linear functions of $1/T$, where T is the mean absolute temperature of the specimen. The graphs for log σ and log $|R|$ against $1/T$ each consist of two rectilinear portions, one covering the high, the other the low temperature results. Thermo-electric measurements yielded results only for the high temperature part.

The theory of impurity semiconductors is developed and found adequate to explain the variation, at any rate at higher temperatures. It is found that an activation energy of about 2 ev. at high temperatures, and from 0·3 to 0·64 ev. at low temperatures, is required for excitation of electrons to the conduction band. The concentration of impurity centres operative in the high temperature range is about $8{\cdot}6 \times 10^{20}$, the electrons having a mean free path of about 10^{-6} cm. The concentration of free " carriers " varies from 10^{11} to 10^{13} as the temperature rises. NiO behaves like a defect semiconductor, " hole " conduction predominating. The significance of these results is discussed.